الدكتور عماد الزغول

الإحصاء التربوي



المحتويات

الفصل الثالث

المئين والرتبة المئينية

الفصل الرابع

مقاييس النزعة المركزية

الفصل الخامس

مقاييس التشتت

الفصل السادس

التوزيع الطبيعي والعلامات المعيارية

الفصل السابع

معاملات الارتباط

الفصل الثامن

التوزيع العيني والتقدير الإحصائي

الفصل التاسع

اختبار الفرضيات الاحصائية

الملاحق

المقدمة

يشهد هذا العصر الذي نعيشه تغيراً واضحاً وملموساً في كافة ميادين الحياة ولاسيما اننا نعيش طفرة الانفتاح والتفجر المعرفي والتفوق التكنولوجي والتي زادت من حجم المطالب والآعباء الإنسانية . إن مثل هذه التغيرات المتسارعة مما لايدعو للشك اصبحت تتطلب التخطيط الحذر والمستمر للتعامل معها ومواجهة المستقبل . ومن هنا ، نجد آن الاهتمام في استخدام ا لمناهج والاساليب الإحصائية المختلفة وتوظيفها في دراسة وتحليل الخصائص المرتبطة بمختلف أنواع المتغيرات والعوامل بات يشكل عنصراً هاماً في كافة الميادين : الاجتماعية والنفسية والتربوية والصناعية والاقتصادية والعلمية وغيرها من الميادين الأخرى .

وبالرغم من تباين الآساليب الإحصائية وتعدد اغراضها واستخداماتها ، إلا آنها في مجموعها تساعد في وصف البيانات الكبيرة المتعلقة بالتغيرات الإحصائية وفي عملية تحليلها ، الأمر الذي يساعد في عمل الاستنتاجات واصدار الاحكام واتخاذ القرارات المناسبة التي تساعد في عمليات التخطيط .

لقد جاءت مادة هذا الكتاب لتركز على المفاهيم والمبادئ الآساسية في مجال الإحصاء الوصفي بالوقت الذي تم التعرض فيه إلى بعض مبادئ الإحصاء الاستدلالي .

وقد تمثل الهدف الآساسي من وراء هذا الكتاب في خدمة غرضين أساسيين هما :

أ- تزويد القارئ بالمعارف حول المبادئ والمفاهيم الإحصائية المتعددة واغراضها واستخداماتها المتعددة .

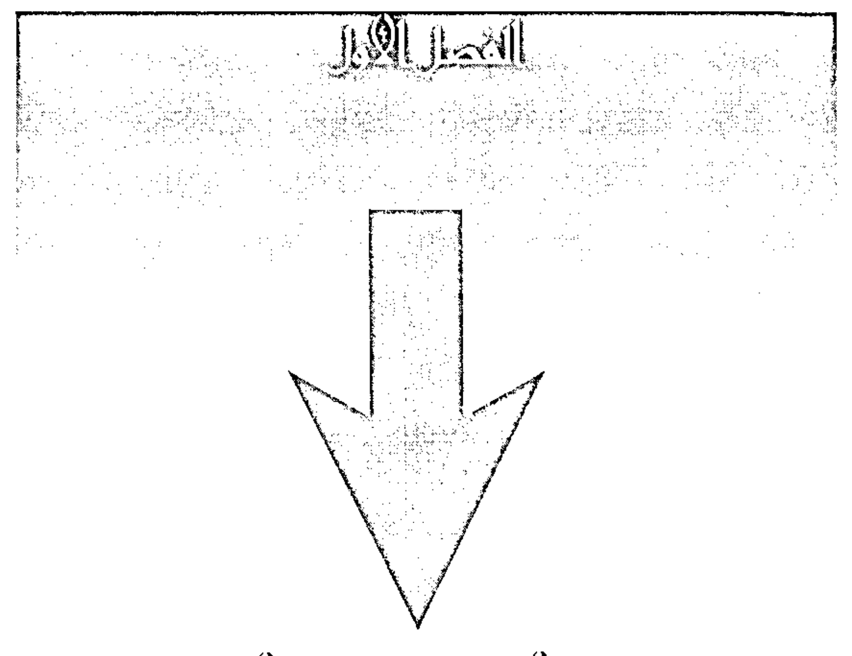
ب- تنمية مهارات القارئ في كيفية حساب وايجاد مثل هذه المبادئ والمفاهيم وكيفية تفسيرها .

وهكذا نجد أن مادة الكتاب تشتمل على عدد من الفصول تتناول مواضيع مثل المفاهيم الاساسية في الإحصاء وطرق عرض وتمثيل البيانات وتلخيصها والمئينات ومقاييس النزعة المركزية ومقاييس التشتت ومفهوم الدرجة المعيارية والتوزيع المعياري ومعاملات الارتباط والانحدار والتنبؤ الإحصائي واختبار الفرضيات وفترات الثقة .

لقد حاولنا تنظيم محتويات هذا الكتاب على نحو متسلسل من حيث عرض موضوعاته المختلفة وذلك على نحو يساعد القارئ على الإلمام بالمفاهيم الإحصائية المختلفة وكيفية حسابها وتفسير الدلالات المرتبطة بها . واخيراً يبقى هذا الكتاب جهداً إنسانياً لايخلو من العيب وقد نكون ممتنين لتزويدنا بأية معلومات أو ملاحظات من شأنها أن تساعد في إثراءه وإخراجه بصورة آفضل .

والله ولي التوفيق

المؤلف



المفاهيم الأساسية لعلم الإحصاء

Basic Concepts of Statistics

تمهيد:

يحظى الاحصاء باهتمام المختصين والباحثين في العديد من ميادين المعرفة المختلفة؛ فهو يعد أحد أكثر فروع المعرفة تداخلاً مع الفروع العلمية الأخرى وأكثرها انتشاراً في المجالات الحياتية المتعددة. ويمكن عزو ذلك الى ان الاحصاء بحد ذاته لا يشكل علماً أوجد لذاته او خدمة اغراضه الخاصة وذلك كما هو الحال في فروع المعرفة الأخرى مثل الفيزياء او الهندسة او الطب او الاحياء او الكيمياء او حتى العلوم الانسانية والاجتماعية الأخرى كعلم النفس والاقتصاد والاجتماع، وانما تتمثل أهميته في ما يقدمه من مناهج واساليب إحصائية يمكن توظيفها بشكل فاعل في مختلف ميادين المعرفة، مما يساعدها بالتالي في تحقيق اغراضها.

فالاحصاء هو بمثابة احد فروع المعرفة الخدماتية تتمثل مهمته الاولى في توليد المعرفة والاساليب والمناهج الاحصائية المتعددة التي تستخدم لاستخلاص ومعالجة البيانات الرقمية المتعلقة بالظواهر العلمية والنفسية والاجتماعية. ومن هنا نجد ان مختلف فروع المعرفة العلمية منها مثل الفيزياء والكيمياء والهندسة والطب والفلك والاحياء وكذلك الانسانية منها مثل علم النفس وعلم الاجتماع وعلوم السياسة والاقتصاد والعلوم العسكرية وعلوم الصناعة والتجارة وكذلك التربية والتعليم توظف الاساليب والمناهج الاحصائية بهدف الوصول إلى معلومات وصفيه او عمل بعض الاستدلالات والاستنتاجات حول العديد من الظواهر موضع البحث، الأمر الذي يساعد في فهم وتفسير مثل هذه الظواهر والتنبؤ والتحكم بها.

وهكذا نجد ان استخدام المناهج والاساليب الاحصائية على المستوى النظري يكاد يكون عديم النفع، اذ إنه لا يقدم أو يؤخر في المجال العلمي. ولكن يصبح لمثل هذه الاساليب قيمة كبرى عندما توظف على البيانات المستمدة من العلوم الاخرى سالفة الذكر.

لقد استخدم الانسان منذ القدم الاحصاء في تنظيم بعض الجوانب الحياتية. وقد كان مفهوم الاحصاء انذاك موازياً لمفهوم العد، اذ اقتصر استخدامه على توظيف بعض الاساليب الاحصائية البدائية

من اجل خدمة بعض الاغراض. وتشير الشواهد ان قدماء الفراعنة المصريين والاغريق اليونانيين استخدموا بعض مبادئ الاحصاء مثل عد او احصاء الموارد المتاحة وحصر عدد الجنود وعدد السكان من اجل تنظيم الميزانية وترتيب الخطط اوقات الحرب والسلم، وبذلك فقد اقتصر استخدامه لخدمة اغراض الدولة وهذا يفسر سبب التسمية اللاتينية له بـ "Statistics" والمشتقة من كلمة (State) والتي تعني « الدولة».

ونتيجة لتطور مناحي الحياة المختلفة وازدياد المطالب واعداد السكان وارتفاع نسب المواليد والوفيات وزيادة علاقات الدول الخارجية وتعدد مصادر التبادل التجاري وعمليات التسويق وزيادة مصادر الموارد وارتفاع حجم الانتاج الزراعي والصناعي اصبحت الحاجة ماسة لاستخدام اساليب احصائية اكثر تعقيداً، اذ ان الاساليب الاحصائية البدائية التي تستند الى العد لم تعد مناسبة لتحقيق اغراض الدولة، ولا سيما انه في الكثير من الحالات يتطلب الامر القفز وراء الاعداد او الارقام المتوفرة وذلك من اجل عمل بعض الاستنتاجات او اصدار الاحكام التي من شأنها ان تسهم في التخطيط لمطالب الحاضر والمستقبل، وبذلك نجد أن تطوراً حدث على الاساليب الاحصائية نتج عنه توليد مبادئ ومناهج احصائية جديدة أكثر ملائمة وتعقيداً.

إن ثمة امراً آخر ساعد في تطويرالمبادئ والاساليب الاحصائية تمثل في انفصال العلوم عن الفلسفة، ودعوة العلماء الى استخدام الطرق العلمية في دراسة الظواهر المختلفة. ومن المعروف ان الطريقة العلمية تستند الى درجة كبيرة الى استخدام الاساليب والطرق الاحصائية البسيطة والمعقدة منها، وهذا ادى بالتالي الى البحث في تطوير الاساليب الاحصائية وابتكار العديد من المبادئ والمناهج الاحصائية الجديدة.

وحديثاً وكنتيجة للتطور العلمي والتكنولوجي في مجال الحاسوب الالكتروني فقد تم استحداث العديد من البرامج الاحصائية المحسوبة التي تمكّن من اجراء العمليات الاحصائية المعقدة على كم هائل من البيانات وفي وقت قصير جداً، مما جعل من استخدام العديد من الاساليب والمناهج الاحصائية امراً في غاية السهولة. ومن الرزم أو البرامج الاحصائية المعروفة الآن والأكثر شيوعاً هي: SPSS NCSS, Dmap, Sas, Minitab, Systat, Statistica.

تعريف علم الإحصاء:

يعنى الاحصاء بالدرجة الاولى في عمليات جمع البيانات الرقمية حول خصائص الاشياء والعمل على تلخيصها وتحليلها وتفسيرها بهدف الوصول الى قرارات او نتائج معينة حول المجتمع الاحصائي

الذي تم أخذ البيانات الاحصائية منه . فعلم الاحصاء هو بمثابة علم البيانات "Science of data" والذي يشتمل على شريحة واسعة من المبادئ والاساليب التي يمكن بواسطتها تلخيص البيانات في صيغ رقمية على نحو يسهل عملية معالجتها للوصول الى استنتاجات او احكام محددة (ماضي وعثمان، 1999). وهكذا فيمكن تعريف الاحصاء على انه علم البيانات الذي يتضمن عمليات جمع وتلخيص وتبويب وتنظيم وتحليل وتفسير البيانات بهدف الوصول الى استنتاجات او احكام معينة (Sincich, 1993).

الأساليب الإحصائية:

يمكن تقسيم الاساليب الاحصائية الى مجموعتين اعتماداً على الهدف الذي من اجله تستخدم ، وتتمثل هذه الاساليب في الآتي :

أولاً: اساليب الاحصاء الوصفي Descriptive Statistics : يستخدم الاحصاء الوصفي عندما يكون الهدف الاساسي هو وصف البيانات الاحصائية المأخوذة من مجتمع ما . وفي الغالب فإن هذه الاساليب تستخدم قاعدة البيانات المتوفرة في المجتمع جميعها وتعمل على تبويبها وتلخيصها بطريقة تساعد في عملية وصفها (Shristensen & Stoup, 1986).

ومثل هذه الاساليب تعمل على تبويب وتلخيص البيانات في جداول تكرارية وتمثيلها بأشكال بيانية معينة والعمل على حساب بعض الاحصائيات مثل الوسط والوسيط والمنوال والمئينات والنسب المئوية وغيرها . وهكذا فالاحصاء الوصفي يعرّف على أنه ذلك الفرع الذي يعنى بعمليات جمع البيانات وتنظيمها وتلخيصها بهدف تقديم اوصاف محددة حول هذه البيانات.

ثانياً: اساليب الاحصاء الاستدلالي Inferential Statistics : يشتمل الاحصاء الاستدلالي على مجموعة من الطرائق والاساليب التي تمكننا من عمل بعض الاستنتاجات والاستدلالات حول خصائص مجتمع احصائي معين من خلال استخدام عينة جزئية من ذلك المجتمع . خلافاً لاساليب الاحصاء الوصفي ، فأن أساليب الاحصاء الاستدلالي تعتمد على الاسلوب العيني المتمثل في أخذ مجموعة صغيرة من مجتمع معين ودراسة خصائص هذه المجموعة بهدف اصدار تعميمات حول المجتمع الذي اخذت منه هذه العينة . وعليه فالاحصاء الاستدلالي هو ذلك الفرع الذي يهتم بدراسة خصائص عينة جزئية من البيانات من اجل عمل بعض الاستدلالات حول خصائص المجتمع الكلي الذي اخذت منه تلك العينة (Sincich, 1993).

من اجل خدمة بعض الاغراض. وتشير الشواهد ان قدماء الفراعنة المصريين والاغريق اليونانيين استخدموا بعض مبادئ الاحصاء مثل عد او احصاء الموارد المتاحة وحصر عدد الجنود وعدد السكان من اجل تنظيم الميزانية وترتيب الخطط اوقات الحرب والسلم، وبذلك فقد اقتصر استخدامه لخدمة اغراض الدولة وهذا يفسر سبب التسمية اللاتينية له بـ "*Statistics*" والمشتقة من كلمة (*State*) والتي تعني « الدولة».

ونتيجة لتطور مناحي الحياة المختلفة وازدياد المطالب واعداد السكان وارتفاع نسب المواليد والوفيات وزيادة علاقات الدول الخارجية وتعدد مصادر التبادل التجاري وعمليات التسويق وزيادة مصادر الموارد وارتفاع حجم الانتاج الزراعي والصناعي اصبحت الحاجة ماسة لاستخدام اساليب احصائية اكثر تعقيداً، اذ ان الاساليب الاحصائية البدائية التي تستند الى العد لم تعد مناسبة لتحقيق اغراض الدولة، ولا سيما انه في الكثير من الحالات يتطلب الامر القفز وراء الاعداد او الارقام المتوفرة وذلك من اجل عمل بعض الاستنتاجات او اصدار الاحكام التي من شأنها ان تسهم في التخطيط لمطالب الحاضر والمستقبل ، وبذلك نجد أن تطوراً حدث على الاساليب الاحصائية نتج عنه توليد مبادئ ومناهج احصائية جديدة أكثر ملائمة وتعقيداً.

إن ثمة امراً آخر ساعد في تطوير المبادئ والاساليب الاحصائية تمثل في انفصال العلوم عن الفلسفة، ودعوة العلماء الى استخدام الطرق العلمية في دراسة الظواهر المختلفة. ومن المعروف ان الطريقة العلمية تستند الى درجة كبيرة الى استخدام الاساليب والطرق الاحصائية البسيطة والمعقدة منها، وهذا ادى بالتالي الى البحث في تطوير الاساليب الاحصائية وابتكار العديد من المبادئ والمناهج الاحصائية الجديدة.

وحديثاً وكنتيجة للتطور العلمي والتكنولوجي في مجال الحاسوب الالكتروني فقد تم استحداث العديد من البرامج الاحصائية المحسوبة التي تمكّن من اجراء العمليات الاحصائية المعقدة على كم هائل من البيانات وفي وقت قصير جداً، مما جعل من استخدام العديد من الاساليب والمناهج الاحصائية امراً في غاية السهولة. ومن الرزم أو البرامج الاحصائية المعروفة الآن والأكثر شيوعاً هي: SPSS NCSS, Dmap, Sas, Minitab, Systat, Statistica.

تعريف علم الإحصاء:

يعنى الاحصاء بالدرجة الاولى في عمليات جمع البيانات الرقمية حول خصائص الاشياء والعمل على تلخيصها وتحليلها وتفسيرها بهدف الوصول الى قرارات او نتائج معينة حول المجتمع الاحصائي

الذي تم أخذ البيانات الاحصائية منه. فعلم الاحصاء هو بمثابة علم البيانات "Science of data" والذي يشتمل على شريحة واسعة من المبادئ والاساليب التي يمكن بواسطتها تلخيص البيانات في صيغ رقمية على نحو يسهل عملية معالجتها للوصول الى استنتاجات او احكام محددة (ماضي وعثمان، 1999). وهكذا فيمكن تعريف الاحصاء على انه علم البيانات الذي يتضمن عمليات جمع وتلخيص وتبويب وتنظيم وتحليل وتفسير البيانات بهدف الوصول الى استنتاجات او احكام معينة (Sincich, 1993).

الأساليب الإحصائية:

يمكن تقسيم الاساليب الاحصائية الى مجموعتين اعتماداً على الهدف الذي من اجله تستخدم، وتتمثل هذه الاساليب في الآتي:

أولاً: اساليب الاحصاء الوصفي Descriptive Statistics: يستخدم الاحصاء الوصفي عندما يكون الهدف الاساسي هو وصف البيانات الاحصائية المأخوذة من مجتمع ما. وفي الغالب فإن هذه الاساليب تستخدم قاعدة البيانات المتوفرة في المجتمع جميعها وتعمل على تبويبها وتلخيصها بطريقة تساعد في عملية وصفها (Shristensen & Stoup, 1986).

ومثل هذه الاساليب تعمل على تبويب وتلخيص البيانات في جداول تكرارية وتمثيلها بأشكال بيانية معينة والعمل على حساب بعض الاحصائيات مثل الوسط والوسيط والمنوال والمئينات والنسب المئوية وغيرها. وهكذا فالاحصاء الوصفي يعرّف على أنه ذلك الفرع الذي يعنى بعمليات جمع البيانات وتنظيمها وتلخيصها بهدف تقديم اوصاف محددة حول هذه البيانات.

ثانياً: اساليب الاحصاء الاستدلالي Inferential Statistics: يشتمل الاحصاء الاستدلالي على مجموعة من الطرائق والاساليب التي تمكننا من عمل بعض الاستنتاجات والاستدلالات حول خصائص مجتمع احصائي معين من خلال استخدام عينة جزئية من ذلك المجتمع. خلافاً لاساليب الاحصاء الوصفي، فأن أساليب الاحصاء الاستدلالي تعتمد على الاسلوب العيني المتمثل في أخذ مجموعة صغيرة من مجتمع معين ودراسة خصائص هذه المجموعة بهدف اصدار تعميمات حول المجتمع الذي اخذت منه هذه العينة. وعليه فالاحصاء الاستدلالي هو ذلك الفرع الذي يهتم بدراسة خصائص عينة جزئية من البيانات من اجل عمل بعض الاستدلالات حول خصائص المجتمع الكلي الذي اخذت منه تلك العينة (Sincich, 1993).

المجتمع الإحصائي Statistical Population

من المعروف ان البيانات التي يتم جمعها هي في الغالب بيانات رقمية تعكس مقادير كمية او خصائص نوعية تتعلق عادة بخصائص معينة لمجموعة من العناصر او الاشياء. وتسمى مجموعة الاشياء او العناصر او الوحدات بالمجتمع الاحصائي. وعليه فالمجتمع الاحصائي هو المجموعة الكلية من العناصر او الافراد او الوحدات او العناصر التي يتم جمع البيانات الاحصائية عن خصائصها. فعلى سبيل المثال، اذا كان الهدف هو جمع بيانات احصائية عن الخبرات العلمية لمعلمي مادة الاجتماعيات في الاردن، فعندها يشكل جميع معلمي مادة الاجتماعيات في الاردن مجتمع الدراسة او المجتمع الاحصائي. اما اذا كان الهدف فحص خلايا جسم شخص ما، فعندها تشكل جميع الخلايا الجسمية لذلك الشخص المجتمع الاحصائي. في حين، اذا كان الهدف هو دراسة حجم إنتاج اشجار الزيتون في منطقة الكرك، عندها تشكل كافة اشجار الزيتون في تلك المحافظة مجتمع الدراسة.

تعرف العناصر او الوحدات التي يتألف منها المجتمع الاحصائي بالوحدات التجريبية "Experimental units" حيث يتم جمع البيانات الاحصائية عن خصائص هذه الوحدات. فعلى سبيل المثال، عند دراسة اطوال طلاب مدرسة معينة، فعندها كل طالب في هذه المدرسة هو بمثابة وحدة تجريبية تشتمل عليها هذه المدرسة بحيث يصار الى قياس الطول لها. وهكذا فالوحدات التجريبية هي العناصر المكونة للمجتمع الاحصائي موضع الاهتمام (Sincich , 1993).

يختلف حجم المجتمع الاحصائي باختلاف عدد العناصر التي يتألف منها او الهدف من جمع البيانات. فقد يكون المجتمع كبيراً جداً كما هو الحال في عدد سكان دولة او اقليم معين، او كما هو في حال دراسة اشجار الزيتون مثلاً في الاردن. وقد يكون صغيراً جداً مثل دراسة عدد الاطباء في مستشفى معين أو عمداء الكليات في جامعة ما.

تسمى القيم الاحصائية التي يتم التوصل اليها عند دراسة خصائص معينة للمجتمع الاحصائي بالمعالم "Parameters". فعند حساب الوسط الحسابي او المنوال او الوسيط او الانحراف المعياري لخصائص المجتمع الاحصائي الكلي، فعندها تعد مثل هذه القيم معالم احصائية لذلك المجتمع، ومثل هذه الخصائص تعتبر دقيقة تماماً، وتعبر بشكل جيد عن خصائص ذلك المجتمع في ظل الظروف التي تم دراسة ذلك المجتمع فيها.

العينة الاحصائية Statistical Sample

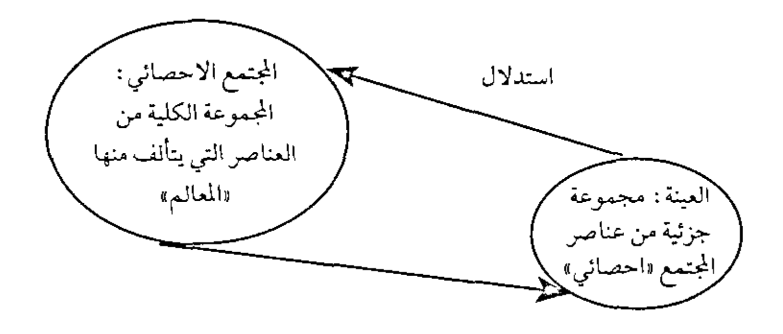
في الكثير من الحالات يصعب جمع البيانات الاحصائية عن خصائص مجتمع ما بسبب كون هذا المجتمع كبير اجداً او غير محدد او لاسباب فنية أو ادارية او مادية. ففي مثل هذه الحالات، يتم تقدير خصائص هذا المجتمع من خلال دراسة خصائص مجموعة جزئية من ذلك المجتمع تعرف باسم العينة "Sample".

فمفهوم العينة يشير الى المجموعة الجزئية من الوحدات او العناصر التي يتم اخذها بطريقة معينة من مجتمع احصائي ما بهدف دراسة خصائصها ، وذلك ليصار تقدير خصائص المجتمع الكلي من خلالها .

تسمى القيم الاحصائية التي يتم حسابها من خلال بيانات العينة بالاحصائيات "Statistics" ومثل هذه القيم تستخدم لتقدير معالم (خصائص) المجتمع الذي اخذت منه العينة (& Christensen Stoup,1986) . ولضمان الدقة في تقدير معالم المجتمع ، يجب ان تكون العينة ممثلة للمجتمع ، وتكون العينة ممثلة عندما تتوزع فيها الخصائص الموجودة بالمجتمع بنفس النسب او بنسب مقاربة .

يتم الحصول على عينة ممثله من خلال اجراء احصائى يعرف بالاختيار العشوائى لافراد العينة "Random Selection" . ففي هذا الاجراء تكون احتمالية اختيار جميع افراد المجتمع متساوية ؛ إذ إن احتمالية اختيار أي فرد من افراد المجتمع ليكون احد افراد العينة تكون مماثلة لاحتمالية اختيار اي فرد اخر في ذلك المجتع . وبالرموز يمكن الاشارة لهذه الاحتمالية على النحو الاتي (ح $= \frac{1}{ن}$) ، حيث (ح) هي احتمال الاختيار لاي فرد من افراد المجتمع ، في حين تمثل (ن) عدد افراد المجتمع الكلي .

احصائيا لا يوجد اتفاق محدد حول الحجم الذي يفترض ان تتألف منه العينة حتى تعطي تمثيلاً حقيقياً للمجتمع الذي تؤخذ منه ، إذ إن بعض الباحثين يرى ان حجم العينة المناسب ليكون ممثلا للمجتمع يجب ان لا يقل عن 5% إلى10% من حجم المجتمع . ولكن تشير نتائج الابحاث والدراسات الى ان الحجم المناسب للعينة يجب ان لا يقل عن"30" عنصراً ، وبذلك يمكن القول إن حجم العينة الذي يساوي (30) فأكثر قد يعطي مؤشراً تمثيلياً واضحاً لخصائص المجتمع . هذا ويوضح الشكل (1:1) العلاقة بين المجتمع والعينة الاحصائية .



شكل (1:1) العلاقة بين المجتمع والعينة الاحصائية

❖ لاحظ في الشكل اعلاه ، ان العينة هي مجموعة جزئية تؤخذ من المجتمع ، بحيث يتم معالجة البيانات التي توفرها العينة لحساب بعض الاحصائيات التي تستخدم لعمل استدلالات حول معالم المجتمع الاحصائي (النبهان ، 2001).

أنواع العينات Types of Samples

تقسم العينات الاحصائية الى مجموعتين هي : العينات الاحتمالية والعينات غير الاحتمالية :

أولاً : العينات الاحتمالية:

تعرف هذه العينات باسم العينات العشوائية "Random Samples" وتمتاز بوجود فرص متكافئة لجميع افراد المجتمع لكي يكونوا احد افراد العينة . بمعنى أن هناك فرصة متكافئة لكل فرد في المجتمع ليكون احد افراد العينة ، وهي مساوية لفرص اختيار الافراد الآخرين في ذلك المجتمع . تستخدم هذه الانواع من العينات عندما يكون المجتمع الاحصائي معروفاً ومحدداً بحيث يتم اختيار افراد العينة بطريقة غير انتقائية ؛ أي على نحو عشوائي وفق شروط محددة تتضمن نوع العينة المطلوب ومدى التجانس والتباين في المجتمع . تشمل هذه المجموعة انواعاً مختلفة من العينات تتمثل بالآتي :

1- **العينة العشوائية البسيطة** Simple Random Sample

تعد هذه الطريقة من افضل الانواع لاختيار عينة ممثلة وذلك لانها تكفل تساوي فرص الاختيار لجميع افراد المجتمع بأن يكونوا احد عناصر العينة . ويمكن استخدام هذا النوع من العينات في حالة كون المجتمعات الاحصائية معروفة ومحددة وحجمها قليل وبالامكان حصره بسهولة . ويتم اختيار افراد العينة في هذا النوع من خلال استخدام القرعة أو ما يعرف بطريقة السحب من الطاقية "Hat way" او من خلال جداول الارقام العشوائية (ملحق رقم 1). عموماً يتطلب استخدام هذه الطريقة اتباع الخطوات التالية :

أ- تحديد عناصر المجتمع (حجم المجتمع).

ب- اعطاء ارقام او رموز معينة حسب تسلسل معين لعناصر المجتمع .

جـ- الاختيار على نحو عشوائي من بين هذه الارقام مع ارجاع الرقم الذي يتم اختياره وذلك لضمان تساوي فرص الاختيار لجميع افراد المجتمع .

فعلى سبيل المثال ، اذا كان لدينا مجتمع من الطلاب يتألف من (100) طالب وكان الهدف هو اختيار عينة ممثلة من هذه المجتمع عدد افرادها (10) ، فعندها يمكن اللجوء الى أحد الاساليب التالية لاختيار افراد هذه العينة .

1-اعطاء ارقام متسلسلة لافراد هذا المجتمع على النحو التالي 003,002,001 ,.......100 . ثم اختيار ارقاما من بين هذه الارقام على نحو عشوائي ، بحيث يمثل الافراد الذين يحملون هذه الارقام افراد العينة وذلك كما هو موضح بالآتي : لنفرض ان الارقام التي تم اختيارها هي 96,72,69,41,37,28,21,13,8,3, فعندها يكون الاشخاص الذين يحملون هذه الارقام هم افراد العينة .

2-يمكن كتابة اسماء الطلاب على قصاصات صغيرة من الاوراق ثم دمجها معاً في وعاء بحيث يتم سحب ورقة في كل مرة ويكون الشخص المكتوب اسمه فيها احد افراد العينة ، ويعاد ارجاع هذه الورقة مرة أخرى الى الوعاء ويتم سحب ورقة أخرى وهكذا الى ان يتم اختيار عشرة افراد .

بعد اعطاء الافراد ارقاماً متسلسلة كما هو الحال في الاجراء الاول ، يمكن اللجوء الى جدول الارقام العشوائية ، بحيث يتم اختيار نقطة بداية على نحو عشوائي في هذا الجدول ، ثم قراءة الارقام في الجدول اما على نحو عمودي او افقي .

ففي حالة مثالنا السابق يجب قراءة الارقام في اول ثلاث منازل في الجدول بحيث يتم اختيار الارقام التي تساوي او تقل عن (100) ليشكل حاملو هذه الارقام عناصر العينة . ولتوضيح هذا الاجراء يمكن النظر الى المثال التالي :

67032	29100	00846	35031	65178
12722	36015	97011	13532	24032
18935	13437	85067	98197	65146
79985	96049	32028	76097	18053
43663	31039	46239	17056	99187
66751	53816	87511	34047	76543

لنفرض انه تم اختيار الرقم المشار اليه بالدائرة كنقطة بداية ، نلاحظ ان الفرد الذي يحمل رقم (31) سوف يكون احد افراد العينة ، فإذا اتخذنا الاتجاه الافقي نلاحظ ان الارقام التي تحتها خط سوف يتم اختيار الافراد الذين يحملونها ليكونوا افراداً بالعينة ، اما اذا اتجهنا على نحو عمودي فعندها نلاحظ ان الارقام التي تحتها خط سوف يمثل افرادها عناصر العينة . وهكذا يتم تكرار هذا الاجراء الى ان يتم تأمين حجم العينة المطلوب .

3- هناك اساليب اخرى تتمثل في استخدام الدواليب او الكرات المرقمة التي تنفخ بالهواء عبر وعاء معين او بعض البرامج الحاسوبية يمكن من خلالها اختيار عينة عشوائية بسيطة .

2- العينة المنتظمةSystematic Sample

يشبه هذا الاسلوب العيني الى حد كبير اسلوب العينة العشوائية البسيطة مع وجود فارق بسيط بينهما يتمثل في ثبات الفاصل العددي الذي يفصل بين الارقام التي يتم اختيارها، وهو ما يعرف بطول فترة الاختيار . وتحديداً يتطلب هذا الاسلوب اعطاء ارقاما متسلسلة لافراد المجتمع، ثم قسمه عدد عناصر المجتمع على عدد العينة المطلوب للحصول على طول فترة الاختيار (الفاصل العددي) ، ثم يصار على نحو عشوائي اختيار رقماً اقل من طول فترة الاختيار ليمثل نقطة البداية بحيث يتم اضافة طول فترة الاختيار مرة أخرى للرقم الجديد للحصول على الرقم الثالث، وهكذا الى ان يتم اختيار حجم العينة المطلوب. ولتوضيح ذلك فلنأخذ المثال التالي: لنفرض ان لدينا مجتمعاً يتألف من (150) فرداً والمطلوب هو اختيار عينة عدد افرادها (15) فرداً.

الحل:

اولاً: نحدد طول فترة الاختيار من خلال قسمة حجم المجتمع على حجم العينة المطلوب (150÷15=10) وعليه فان الرقم (10) يمثل طول فترة الاختيار.

ثانياً: يتم اختيار رقما على نحو عشوائي اقل من طول فترة الاختيار ليكون نقطة البداية. ولنفرض انه تم اختيار رقم (7)، عندها حامل الرقم (7) يكون احد افراد العينة، اما الفرد الثاني في العينة هو الذي يحمل الرقم (17) اي (7+10) ، اما الثالث فهو الذي حمل رقم (27؛ اي 17+10)، اما الرابع فذلك الذي يحمل الرقم (37) . وعموماً فإن افراد العينة في مثالنا السابق هم الأفراد الذين يحملون الارقام التالية (147,137,127,117,107,97,87,77,67,57,47,37,27,17,7).

3-العينة الطبقية "Striatified Sample"

يستخدم هذا النوع من العينات في حالة كون مجتمع الدراسة غير متجانسا ويتألف من فئات او طبقات متباينة، فعندها يتطلب الامر تحديد حجم هذا المجتمع، ومن ثم تحديد الفئات او الطبقات التي يتألف منها وحجم كل طبقة أو فئة من فئاته، بحيث يصار الى اختيار عددا من الافراد من كل فئة على نحو يتناسب وحجم الفئة بالنسبة للحجم الكلي للمجتمع. وعموماً يمكن اتباع الخطوات التالية في هذا النوع من العينات:

1- تحديد حجم المجتمع الاحصائي وعدد الفئات (الطبقات) التي يتألف منها وحجم كل طبقة.
2- تحديد حجم العينة المطلوب.
3- تحديد عدد الافراد الذين يفترض اختيارهم من كل طبقة وفقاً للمعادلة التالية:

$$\frac{\text{حجم الطبقة}}{\text{حجم المجتمع}} \times \text{حجم العينة}$$

4 - استخدام احدى الطرق العشوائية السابقة لاختيار الافراد في كل طبقة من طبقات المجتمع .

ولتوضيح ذلك فلنأخذ المثال التالي : لنفرض ان لدينا مجتمعا احصائيا (طلبة جامعة ما) عدد افراده (2000) طالباً موزعين على التخصصات المبينة في الجدول ادناه ، والمطلوب اختيار عينة عدد افرادها (200) طالبا من طلاب هذه الجامعة .

التخصص	لغة عربية	لغة انجليزية	حقوق	علوم	تاريخ	هندسة	طب	المجموع
عدد الطلاب	380	200	260	400	240	300	220	2000

وعليه فإن توزيع افراد العينة حسب التخصصات سوف يكون على النحو الآتي :

عدد طلاب اللغة العربية = $\frac{380}{2000} \times 200 = 38$ طالباً

عدد طلاب اللغة الانجليزية = $\frac{200}{2000} \times 200 = 20$ طالباً

عدد طلاب الحقوق = $\frac{260}{2000} \times 200 = 26$ طالباً

عدد طلاب العلوم = $\frac{400}{2000} \times 200 = 40$ طالباً

عدد طلاب التاريخ = $\frac{240}{2000} \times 200 = 24$ طالباً

عدد طلاب الهندسة = $\frac{300}{2000} \times 200 = 30$ طالباً

عدد طلاب الطب = $\frac{220}{2000} \times 200 = 22$ طالباً

مجموع افراد العينة = 200 طالبا

4- **العينة العنقودية** Cluster Sample

يستخدم هذا النوع من العينات عندما يكون حجم المجتمع الاحصائي كبيراً جداً او غير محدد او في حالة وجود صعوبات فنية او ادارية او مادية تحول دون اختيار افراد العينة على المستوى الفردي . ففي

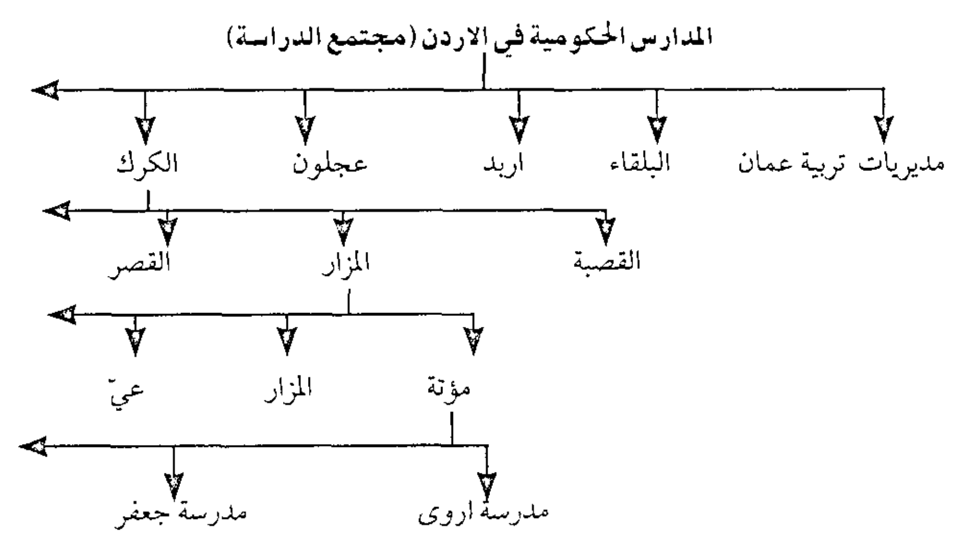
مثل هذا النوع من العينات تكون وحدة الاختيار عنقوداً ، اي مجموعة من العناصر وليس افراداً. وتكمن الفلسفة وراء هذا النوع من العينات في كون ان المجموعة التي تتألف من عدد من العناصر تكفل الى درجة ما تمثيل العينة للمجتمع الاحصائي. فعلى سبيل المثال ، لنفرض ان باحثاً اراد اجراء دراسة على طلبة الصف السادس الاساسي في المدارس الحكومية التابعة لمديرية تربية لواء المزار، فعندها يتطلب الامر من الباحث تحديد جميع المدارس التي تشتمل على الصف السادس الاساسي في لواء المزار، ومن ثم يعمد الى اختيار بعض الصفوف على نحو عشوائي من هذه المدارس لتشكل عينة الدراسة. وهكذا نجد ان الصف الواحد شكل وحدة الاختيار وهو بمثابة عنقود او مجموعة صغيرة.

وتجدر الاشارة هنا، ان في هذا النوع من العينات يفضل تحديد حجم العينة على نحو مسبق لانه يصعب في بعض الحالات الالتزام بهذا الحجم طالما ان وحدة الاختيار هي العنقود وليس الفرد.

5- **العينة متعددة المراحل** Multi- Stage Sample

يصعب في بعض الحالات اللجوء الى الاساليب السابقة لاختيار عينة عشوائية بسبب وجود بعض العوائق مثل ضيق الوقت او ارتفاع حجم التكاليف او زيادة حجم الجهد المطلوب لذلك، او بسبب كون مجتمع الدراسة متشعباً وواسعاً على نحو يجعل من عملية الاحاطة بجميع وحداته امراً صعباً. ففي مثل هذه الحالة، يتم تقسيم المجتمع الى وحدات واختيار احداها عشوائياً، ثم يعمد الباحث الى تقسيم هذه الوحدة الى وحدات اصغر واختيار احداها عشوائياً، ثم تقسيمها الى وحدات اصغر ومن ثم اختيار احداها عشوائياً، وهكذا الى أن يتم الحصول على العينة المطلوبة.

ولتوضيح اجراءات هذا النوع من العينات، يمكن الاستعانة بالمخطط التوضيحي التالي :



لاحظ اعلاه : المطلوب هو اختيار بعض المدارس الحكومية للدراسة ، ولتحقيق هذا الغرض ، تم تقسيم مجتمع الدراسة حسب المديريات التابعة للمحافظات ، ومن ثم جرى اختيار محافظة عشوائيا ، ثم جرى اختيار احدى المدريات التابعة لها على نحو عشوائي ، وبعد ذلك تم اختيار احدى المدن عشوائياً ، ومن المدارس الموجودة في هذه المدينة جرى اختيار بعض المدارس على نحو عشوائي لتشكل عينة الدراسة .

ثانياً: العينات غير الاحتمالية:

تسمى هذه بالعينات غير العشوائية "Non random samples" ، ومثل هذه الانواع من العينات تكون في اغلب الحالات متحيزة لأن عملية اختيارها لم تتم بالطرق الصحيحة . وفي اغلب الحالات يتم اللجوء الى هذا النوع من العينات في حالة عدم القدرة على تحديد المجتمع الاحصائي على نحو دقيق ، او عندما يكون الهدف من الدراسة هو دراسة حالات معينة او عينات محددة . وتشتمل هذه المجموعة على الانواع التالية من العينات :

1- **عينة الصدفة** Accidental Sample : تعرف باسم العينة العرضية بحيث يتم اختيارها على نحو غير اتفاقي ، مثل اخذ بعض البيانات من صحيفة معينة او ملاحظة بعض الحالات والعمل على دراستها والى غير ذلك من الامثلة الاخرى .

2- **العينة القصدية** Purposive sample : يتم اللجوء الى هذا النوع من العينات عندما يكون الهدف دراسة حالات محددة ، حيث يتم اختيار هذه الحالات لتمثل عينة الدراسة . فعلى سبيل المثال ، قد يكون هدف الباحث هو دراسة اوضاع المعلمين في مدرسة معينة ، ففي مثل هذه الحالة ، فإن عينة دراسته تكون مقصودة وتتضمن فقط معلمي هذه المدرسة . كما وقد يلجأ الباحث الى هذا النوع من العينات في حالة وجود معلومات مسبقة بأن هذه العينة تمثل المجتمع الدراسي او عندما يعتقد بأنها قد تكون ممثلة لذلك المجتمع .

3- **عينة المناسبة** Convenient Sample : وهي العينة التي يتم اختيارها في ظل بعض الظروف مثل سهولة الحصول عليها او بسبب قربها من مكان عمل او اقامة الباحث ، كأن يختار الباحث عينة من طلاب المدرسة القريبة من مكان سكناه .

4- **عينة القطعة** Chunk Sample : في هذا النوع من العينات ، يلجأ الباحث الى اقتطاع عدد من افراد المجتمع ليشكلوا عينة الدراسة دون اعتماد اسس الاختيار الصحيح . فعلى سبيل المثال ، قد يلجأ المعلم الى اختيار اول عشرين طالب في كشف علاماته ليمثلوا عينة الدراسة لديه ، اوقد يعمد باحث ما الى مقابلة اول عشرة اشخاص يمرون في شارع معين لاستطلاع آرائهم حول موضوع معين . ويعد هذا النوع من اضعف انواع العينات وهو لا يمثل على الاطلاق مجتمع الدراسة تمثيلاً حقيقياً .

5- **عينة التطوع** Volunteer Sample : يتم اللجوء الى هذا النوع من العينات عندما يكون هناك بعض الصعوبات في الحصول على بعض البيانات من خلال استخدام العينات العشوائية ، اذ يتم في مثل هذه الحالة الاعتماد على بعض الافراد المتطوعين للحصول على البيانات المطلوبة .

6- **العينة الحصصية** Quota Sample : تسمى بعينة الانصبه ، وهي مماثلة للعينة الطبقية ولكنها تختلف عنها من حيث إنها تستخدم في حالة كون مجتمع الدراسة غير محدد . وفي هذا النوع من العينات لا تكون نسب الطبقات محددة كما هو الحال في العينة الطبقية ، وانما يكتفي بأخذ نسب او حصص من كل طبقة موجودة في ذلك المجتمع .

المتغيرات الإحصائية Statistical Variables

من المعروف ان البيانات الاحصائية التي يتم جمعها هي في العادة ترتبط بسمات او خصائص الوحدات التجريبية (العناصر) لمجتمع ما ، ومثل هذه الخصائص (السمات) تسمى بالمتغيرات وذلك لانها تتواجد بنسب مختلفة لدى جميع وحدات او عناصر المجتمع . فعلى سبيل المثال عند قياس اطوال طلاب مدرسة معينة ، نجد جميع الطلاب يمتلكون صفة الطول ولكن سمة الطول هذه تختلف من طالب لآخر حتى عند نفس الفئة العمرية الواحدة . كذلك عند قياس كمية سقوط الامطار في الاردن ، نجد ان كمية السقوط تتباين من منطقة الى أخرى ومن مدينة الى أخرى . ويقاس على ذلك متغيرات اخرى كثيرة مثل مستوى الدافعية ، والذكاء ، واللون ، والجنس ، ودرجات الحرارة والسرعة ، والزمن والمسافة ، والتحصيل ، ولون العيون ، والمستوى التعليمي ، والتخصص ، والمهنة ، وفصيلة الدم والى غير ذلك من المتغيرات الاخرى . وهكذا نلاحظ اننا في المناهج الاحصائية نتعامل دائماً مع متغيرات عشوائية تعرف بالمتغيرات الاحصائية . تصنف المتغيرات العشوائية حسب طبيعتها وطريقة قياسها الى عدة انواع تتمثل بالنحو الآتي :

أولاً : المتغيرات الاسمية Nominal variables : وتعرف ايضاً باسم المتغيرات التصنيفية ، وهي المتغيرات النوعية التي لا يمكن التعبير عنها بمقادير كمية كدلالة على مدى امتلاكها لسمة ما ، ومن هذه المتغيرات الجنس ، والمهنة ، والتخصص وفصيلة الدم واللون ونوع المرض والحالة الاجتماعية والرقم الوطني والرقم الجامعي وارقام اللاعبين الرياضيين . فالارقام اوالقيم العددية التي تعطى لهذه المتغيرات يهدف من وراءها تصنيف هذه المتغيرات فقط بحيث لا تعكس هذه الارقام مقاديراً كمية . فعلى سبيل المثال عند الحديث عن متغير الجنس ، نلاحظ انه يتألف من فئتين هما الذكور والاناث ، فعند اعطاء الذكور رقم (1) والاناث رقم (2) ، فهذا لا يعني ان الاناث افضل من الذكور كون ان الرقم (2) اكبر من الرقم (1) ، وانما الرقم هنا هو مؤشر للفئة التي ينتمي لها كل جنس ، وكما هو الحال في متغير الجنس

يمكن التعامل كذلك مع متغير الحالة الاجتماعية (1) متزوج، (2) مطلق، (3) ارمل، (4) أعزب والمتغيرات الاخرى.

وهكذا نجد ان الارقام اوالمقادير الكمية التي تعطى للمتغيرات وفقاً لهذا المقياس يهدف من وراءها التسمية او التصنيف او التعريف بالمتغيرات فقط. وذلك من اجل تسهيل عملية التعامل مع هذه المتغيرات واجراء بعض العمليات الاحصائية في الحاسوب الالكتروني.

ثانياً: المتغيرات الترتيبية Ordinal Variables: تسمى هذه الفئة من المتغيرات بالمتغيرات الرتبيه وهي تقع على مستوى قياس اعلى من المتغيرات الاسمية وذلك لان الارقام التي تعطى لها تعكس درجات الافضلية بينها (عودة، 1993). فالارقام التي تعطى لمثل هذه المتغيرات تخدم غرضين اساسيين هما:

أ- تصنيف هذه المتغيرات في فئات أو مجموعات تدل عليها.

ب- بيان درجة الافضلية من حيث مدى الامتلاك لسمة معينة، الامر الذي يساعد في ترتيبها تنازلياً اوتصاعدياً.

وتجدر الاشارة هنا، ان الارقام التي تعطى لمثل هذه المتغيرات لا تعكس بالضرورة مقادير كمية، اذ ان المسافات التي تفصل بين الارقام رغم انها تبدو متساوية الا انها لا تعني تساوي الفروق بين المتغيرات. فعلى سبيل المثال، عندما نرتب الافراد حسب الطبقة الاقتصادية ونشير اليها بالارقام كما هو مبين ادناه:

طبقة ذوي الدخل المتدني	1
طبقة ذوي الدخل المتوسط	2
طبقة ذوي الدخل المرتفع (الاغنياء)	3

فنلاحظ هنا، ان الارقام استخدمت لتصنيف الافراد وترتيبهم حسب مستوى الدخل، اي بينت لنا ان افراد الفئة الثالثة (الاغنياء) هم افضل من افراد الفئة الثانية والأولى من حيث الدخل، ولكن بالوقت نفسه لا يعني ان الرقم (3) افضل بمرتين من الرقم (1). وهذا بالتأكيد لا يعني ايضاً ان الفرق بين (2, 3) يساوي بالضرورة الفرق بين (1, 2)؛ اي لا يعني ان الفرق بين دخل افراد الطبقة الغنية والطبقة المتوسطة هو مساوٍ للفرق بين دخل افراد الطبقة المتوسطة والطبقة المتدنية.

وخلاصة القول أن الارقام وفق المقياس الرتبي تعطي للمتغيرات بهدف تصنيفها وترتيبها حسب درجة امتلاكها لسمة معينةفي الوقت الذي لا تعكس الارقام الفروق الكمية الحقيقية بين هذه التغيرات،

وذلك لان الارقام هنا لا تقترن بوحدات حقيقية للقياس . ومن الامثلة على هذا النوع من المتغيرات شدة المرض (شديد، متوسط، متدني) ، المستوى الاقتصادي والاجتماعي (عليا ، متوسطة، متدنية) الرتب العسكرية (عميد، عقيد، مقدم، رائد، نقيب. . .) ، مستوى الكفاءة، ودرجة التعاون وغيرها من المتغيرات الاخرى .

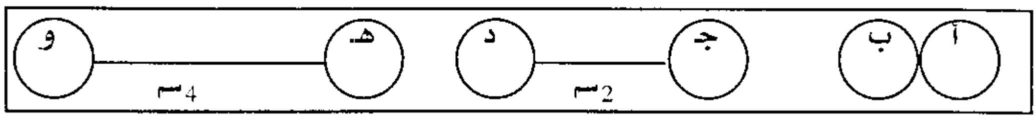
ثالثا: المتغيرات الفئوية Interval Variables : وهي المتغيرات التي تقاس وفق مقياس فئوي ، حيث ان الارقام التي تعطى لمثل هذه المتغيرات تعكس معانٍ كمية من حيث مدى امتلاكها لسمة ما . تعد هذه المتغيرات ارقى من المتغيرات التصنيفية والرتبية ، اذ يمكن المقارنة بينهما على اساس كمي نظراً لتوفر وحدة القياس . وهذا بالطبع يعنى ان المسافات التي تفصل بين الارقام متساوية بحيث تتيح لنا امكانية تحديد الفروق بين المتغيرات واجراء بعض العمليات الحسابية . فعلى سبيل المثال اذا كانت علامات اربعة طلاب على امتحان هي 45, 50, 60, 65، فعندها يمكن القول ان الفرق بين (50,45) مساوٍ للفرق بين (65,60) ، اي ان هذا الفرق مساوٍ لخمس وحدات قياس (النبهان، 2001) .

ونظراً لوجود وحدة قياس ثابتة في حالة هذه المتغيرات ، فهذا يعني توفر خاصية وجود الصفر الافتراضي ، حيث يمثل هذا الصفر قيمة كأي قيمة اخرى يأخذها المتغير . فالصفر الذي تأخذه المتغيرات وفق المقياس الفئوي لا يعني انعدام السمة ؛ اي ان السمة لا توجد لدى المتغير وذلك لان الصفر هنا ليس حقيقياً وانما افتراضياً .

فعند القول على سبيل المثال ، ان درجة حرارة جسم ما تساوي صفراً ، فهذا لا يعني عدم وجود حرارة في ذلك الجسم ، والطالب الذي يحصل على علامة صفر في الامتحان لا يعني انه لا يعرف شيئاً في المادة الدراسية .

هناك العديد من الامثلة على هذا النوع من المتغيرات مثل التحصيل الدراسي ، معامل الذكاء ، مستوى الدافعية ، القلق ، الاكتئاب ودرجات الحرارة والسنوات الدراسية وغيرها من المتغيرات الاخرى .

رابعاً: المتغيرات النسبية Ratio variables : وهي تمثل مجموعة المتغيرات التي تكمم وفق مقياس نسبي في ضوء وجود صفر حقيقي (الصفر المطلق) والذي يعني انعدام وجود السمة او غيابها . وهذا الامر بالتالي يمكن من اجراء جميع العمليات الحسابية على هذه المتغيرات ، اذ يمكن القول بأن المتغير الفلاني يساوي ضعف متغيرآخر او يقل عنه بمعدل النصف مثلاً . ولتوضيح ذلك يمكن النظر الى الشكل (1:2) .



شكل (1:2) يبين متغير المسافة

نلاحظ في الشكل اعلاه ان المسافة بين الدائرتين (أ) و (ب) تساوي صفراً، اي لا توجد مسافة بينهما، وان المسافة بين الدائرتين (هـ) ، (و) تساوي ضعفي المسافة بين الدائرتين (جـ) و (د) .

هناك بعض المتغيرات التي تصنف وفق المقياس النسبي مثل الطول والوزن والكثافة والمسافة والدخل الشهري وغيرها من المتغيرات الأخرى .

بالاضافة الى التصنيف السابق يمكن تصنيف المتغيرات وفق ابعاد اخرى ، وذلك على النحوالاتي : -

أولاً: المتغيرات النوعية(Qualitative Variables):وهي المتغيرات التي تشتمل عدة أصناف مثل الجنس(ذكر، انثى)، الحالة الاجتماعية (متزوج، اعزب، مطلق، ارمل) ، التخصص الاكاديمي (علمي، ادبي، مهني، تحريضي، صناعي) وغيرها . وفي مثل هذه المتغيرات لا يوجد اساس ثابت للمفاضلة بينها او ترتيبها كمياً في ضوء مدى اهميتها نظراً لعدم وجود معيار ثابت للمفاضلة بينها بين زمان وزمان آخر او بين مجتمع ومجتمع آخر . فعلى سبيل المثال ، لا يمكن القول ان الذكور افضل من الاناث او ان التخصص العلمي افضل من المهني وهكذا .

ثانياً: المتغيرات الكمية(Quantitative variables):وهي المتغيرات التي يمكن التعبير عنها كمياً بحيث يمكن المفاضلة بينها في ضوء المقادير الكمية التي تنسب اليها . وتشمل هذه المتغيرات ما يلي :

1 - **المتغيرات الرتبية**:وهي المتغيرات التي يمكن ترتيبها تنازلياً او تصاعدياً في ضوء المقادير الكمية التي تعطى لها . ومن الامثلة عليها الطبقة الاجتماعية، الطبقة الاقتصادية، المرحلة الدراسية، والرتبة العسكرية .

2- **المتغيرات الوثابة** (Discontinuous): وهي تعرف باسم المتغيرات المنفصلة والتي قد تأخذ قيمة ثابتة ومحددة في مدى معين ضمن بعد القياس لها . ففي حالة هذه المتغيرات ليس للكسور او الاجزاء اي معنى ، لان المتغير من هذه المتغيرات يأخذ قيمة صحيحة في مدى معين . فعلى سبيل المثال لو اخذنا اعداد المدارس في محافظات الاردن، وذلك كما هو موضح في الشكل (1:3) .

1800	1300	950	910	780	760	750	742	740	730	610	540
عمان	إربد	الزرقاء	البلقاء	الكرك	العقبة	معان	المفرق	الطفيلة	عجلون	جرش	مأدبا

الشكل(1:3) يوضح مفهوم المتغير المنفصل (عدد المدارس في محافظات الاردن)(*)

* الارقام اعلاه ليست حقيقية وانما يقصد بها توضيح مفهوم المتغير المنفصل .

نلاحظ في شكل (1:3) ان متغير عدد المدارس في محافظات الاردن تم تحديده وفق مدى تراوح بين 540 و1800 مدرسة، بحيث نلاحظ ان عدد المدارس في كل محافظة هو عددصحيح وقد وقع على نقطة محددة ضمن هذا المدى ، بحيث لا يمكن القول مثلاً ان عدد المدارس في محافظة عجلون يساوي530 مدرسة ونصف المدرسة .

3- **المتغيرات المتصلة** (Continuous) : وهي مجموعة المتغيرات التي تأخذ اي قيمة او اجزاء من القيمة ضمن مدى القياس، وغالباً ما تأخذ هذه المتغيرات شكل التوزيع الطبيعي . ومن الامثلة عليها المسافة، والزمن والعمر والتحصيل والذكاء والدافعية والقلق والوزن ومستوى الانتباه وزمن الرجع وغيرها من الامثلة الاخرى . ولتوضيح مفهوم هذه المتغيرات يمكن النظر الى الشكل رقم (1:4) والذي يوضح متغير الوزن .

—xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx—

14.200كغم **شكل(1:4) يوضح مفهوم المتغير المستمر (الوزن)(*)** 16.100كغم

لو نظرنا الى الشكل رقم (1:4) والذي يمثل اوزان (100) طفل تتراوح اعمارهم بين (سنتين) و(ثلاث سنوات) سوف نلاحظ ان اوزانهم سوف تنتشر على جميع نقاط هذا المتصل او المدى وذلك لان المتغير المستمر قد يأخذ قيماً كسرية اوغير صحيحة ضمن مدى القياس الذي يقاس في ضوءه .

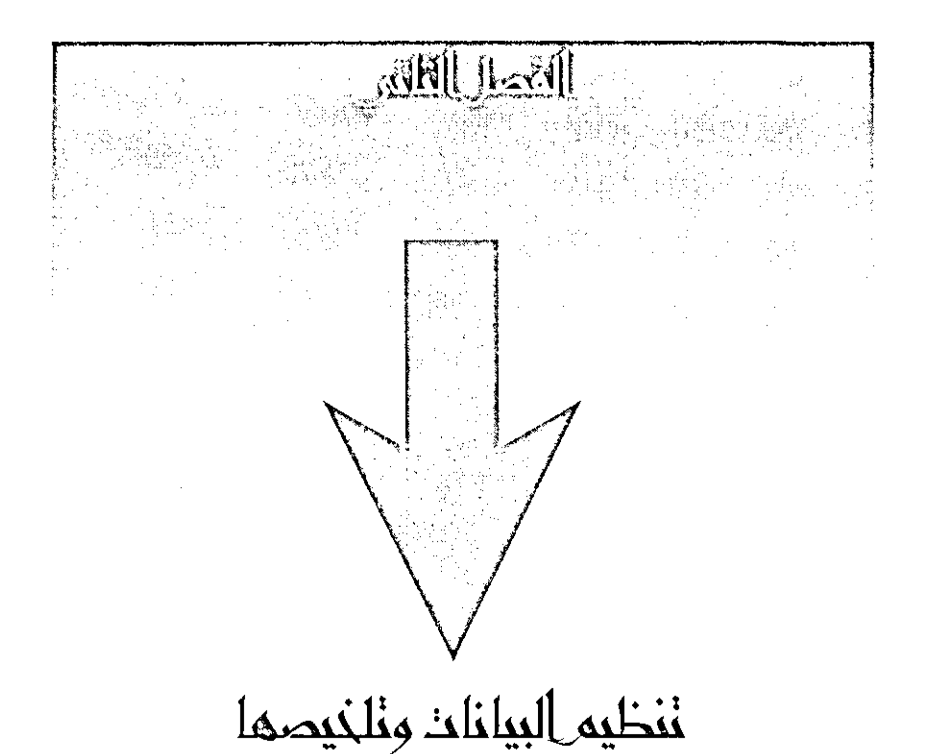
إن ثمة تصنيف آخر للمتغيرات يلجأ اليه الباحثون ولا سيما في الابحاث التجريبية وشبه التجريبية، اذ يتم تصنيف المتغيرات في ضوء اهداف البحث اوالدراسة الى متغيرات مستقلة وتابعة وضابطة . فالمتغير المستقل هو ذلك العامل الذي يختاره الباحث ويتحكم في قيمته لتحديد اثره في متغير آخر يعرف بالمتغير التابع .

اما المتغير التابع فهو الظاهرة موضع الدراسة والتي قد تحدث او لا تحدث، تزيد او تنقص ، تتأثر او لا تتأثر بما يجريه الباحث من معالجات على المتغير المستقل، في حين يمثل المتغير الضابط العامل الذي يسعى الباحث لازالته او التقليل من اثره وذلك من اجل تحديد الاثر الفعلي للمتغير المستقل في المتغير التابع . فعلى سبيل المثال لو أراد باحث الكشف عن اثر عدد ساعات التدريب لدى الطلبة الذكور في درجة اتقان مهارة رياضية معينة، فعندها يكون المتغير المستقل في هذه الدراسة هو عدد ساعات التدريب(**)، اما المتغير التابع فهو اتقان المهارة الرياضية، في حين المتغير او المتغيرات الضابطة فهي الجنس(***) بالاضافة الى عوامل اخرى مثل الوجبات الغذائية، مكان التدريب ، المدرب، العمر، وقت التدريب الخ .

* اشارة (×) تعني الاوزان المحتملة للاطفال ضمن مدى القياس .

** عدد ساعات التدريب : هي المتغير المستقل والتي قد يحاول الباحث التحكم بها من حيث زيادتها او انقاصها او حجبها عن افراد الدراسة لتحديد اثرها في المتغير التابع وهو اتقان المهارة .

*** لاحظ ان الباحث حاول منذ البداية ضبط متغير الجنس، حيث حدد ذلك بإجراء الدراسة على الذكور فقط ومع هذا قد يلجأ الباحث الى ازالة او ضبط عوامل أخرى مثل نوع الغذاء، مكان التدريب، المدرب، العمر، ... الخ .



تنظيم البيانات وتلخيصها

Data Organization and Summerization

تمهيد:

من المعروف أن مجرد جمع البيانات الإحصائية عن متغيرات مجتمع ما يكاد يكون عديم الفائدة، خصوصاً إذا تم التعامل مع هذه البيانات على الصورة التي هي عليه، إذ لا يمكن من خلالها وصف خصائص المجتمع على نحو دقيق أو الوصول إلى أية استنتاجات أو عمل أحكام حيالها. ولا سيما إذا كان حجم البيانات كبيراً جداً، كما هو الحال في معالجة درجات الطلبة المتقدمين لامتحان الشهادة الثانوية أو مبيعات شركة تجارية متعددة الأغراض أو الموارد الاقتصادية المتعددة لدولة ما أو رواتب الموظفين في القطاعات العامة والخاصة وغيرها من الأمثلة الأخرى.

وحتى يتم التعامل مع البيانات الاحصائية الخام التي يتم جمعها عن متغيرات مجتمع ما، يجب العمل على تنظيمها وتلخيصها على نحو يُسَّهل عملية معالجتها وتحليلها. ويمكن تحقيق هذا الغرض من خلال اللجوء إلى التوزيعات التكرارية "Frequency distribution" للبيانات، وذلك من خلال عرض البيانات وتمثيلها في جداول تكراريه أو رسوم بيانية. فالتوزيعات التكرارية هي بمثابة معالجة أولية للبيانات الاحصائية، وقد تشكل هدفاً يتوقف معالجة البيانات عند هذا الحد، أو أنها تشكل خطوة من ضمن خطوات معالجة البيانات وتحليلها بحيث تسهل لاحقاً إجراء المزيد من العمليات الاحصائية المعقدة على هذه البيانات.

يمكن تنظيم البيانات وتلخيصها وفق التوزيعات التكرارية من خلال العرض الجدولي أو التمثيل البياني، بحيث تتباين طريقة العرض تبعاً للهدف الذي من أجله جمعت البيانات ونوعية المتغيرات وعددها. وفيما يلي عرض لطريقة عرض البيانات الاحصائية:

أولاً: العرض الجدولي للبيانات

يعد العرض الجدولي للبيانات خطوة هامة من خطوات معالجة البيانات وتحليلها وذلك لأن الجداول الاحصائية تعمل على تلخيص البيانات الكبيرة وتبويبها وتصنيفها على نحو تبدو ذات معنى وقيمة. فهي تتيح إمكانية تلخيص كم هائل من البيانات وتبويبها وفقاً لعدد من الأبعاد والمتغيرات، الأمر الذي يسهل بالتالي من عمليات معالجتها وتحليلها وإجراء المزيد من العمليات الاحصائية عليها وتحقيق الأغراض المطلوبة.

تختلف الجداول التكرارية "Frequency tables" للبيانات الاحصائية تبعاً لاختلاف طبيعة المتغيرات موضع البحث من حيث كونها كمية أو نوعية وعدد مثل هذه المتغيرات بالإضافة إلى حجم البيانات الاحصائية. هذا وتتباين الجداول التكرارية لتشمل الجداول البسيطة والمعقدة منها. فالجداول التكرارية البسيطة هي التي تلخص البيانات المتعلقة بالمتغيرات النوعية أو أنها تلك التي تلخص بيانات بمتغير كمي واحد. أما الجداول المعقدة فهي التي يتم فيها تناول عدداً من المتغيرات بالوقت نفسه أو أنها تستخدم لتلخص عدد كبير من المفردات أو البيانات الاحصائية، وفيما يلي عرض للجداول التكرارية:

1-الجداول التكرارية البسيطة

وهي الجداول التي تستخدم كقائمة لعرض البيانات الاحصائية المتعلقة بمتغيرات نوعية (تصنيفية) أو بعدد محدود (لا يتجاوز 30 مشاهدة) لمتغير كمي واحد. ومثل هذه الجداول تشتمل على عمودين أو ثلاثة على الأكثر، بحيث يتم تمثيل المتغيرات النوعية في العمود الأول، في حين يتم رصد التكرارات لهذه المتغيرات في العمود الثاني، وقد يستخدم العمود الثالث أحياناً لبيان النسب المئوية. ومن الأمثلة على ذلك كشوف العلامات التي يعدها المعلمون لدرجات طلابهم، أو عدد الإصابات بمرض معين في المحافظات، أو مقدار العائد الاقتصادي لكل قطاع إنتاجي، أو عدد مرات التدريب التي يحتاجها الأفراد لاتقان مهارة رياضة معينة وإلى غير ذلك من الأمثلة الأخرى. ولتوضيح ذلك فلنأخذ الأمثلة التالية: -

مثال (2:1): أمامك قائمة من علامات 30 طالباً على امتحان لمادة الرياضيات موزعة على النحو التالي:

14, 14, 15, 15, 15, 15, 11, 16, 17, 13

10, 10, 11, 10, 18, 22, 22, 30, 26, 25

17, 13, 13, 13, 22, 19, 18, 17, 28, 23

ولتلخيص هذه البيانات يمكن تكوين الجدول التالي:

الدرجة	عدد الطلاب الحاصلين عليها (التكرار)
10	3
11	2
13	4
14	2
15	4
16	1
17	3
18	2
19	1
22	3
23	1
25	1
26	1
28	1
30	1
المجموع	30

يلاحظ في الجدول آعلاه أن البيانات الاحصائية (درجات الطلبة) تم تلخيصها وتصنيفها في جدول يتألف من العمودين: الأول استخدم لعرض الدرجات والآخر لعرض عدد الطلبة الحاصلين عليها.

هذا وقد تتضمن بعض مثل هذه الجداول عاموداً ثالثاً يبين فيه النسب المئوية التي يمثلها كل متغير بالنسبة للمجموع الكلي، وذلك كما هو موضح في المثال (2:2):

مثال(2:2) : يوضح حجم العائد الاقتصادي للقطاعات الإنتاجية ونسبها المئوية

القطاع	حجم العائد	النسبة المئوية
الزراعة	2.323056	%3.91
الصناعة	6.893017	%11.59
التجارة	25.369140	%42.67
البنوك	9.234510	%15.53
حوالات العاملين في الخارج	15.639367	%26.30
المجموع	59.459090	%100

كمـا ويمكن عـرض البـيانات الاحصائيـة في مثـل هذه الجـداول من خـلال اسـتخـدام الإشـارات (Tallys) للدلالة على عدد التكرارات المقابلة لكل حالة، وذلك كما هو الحال في التعامل مع المتغيرات النوعية مثل حصر عدد الأصوات التي يحصل عليها المرشحون لمقعد انتخابي معين. والمثال (2:3) يوضح هذا النوع من الجداول.

مثال (2:3) : تقدم خمسة مرشحين (أحمد، علي، قاسم، عادل، ومنى) لشغل مقعد انتخابي معين، وكان عدد الأصوات التي حصل عليها هؤلاء المرشحون كما هو مبين في الجدول التالي:

اسم المرشح	الاشارات	التكرار
أحمد	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ||	47
قاسم	𝍸 𝍸 𝍸 𝍸 𝍸 ||||	29
عادل	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |	31
منى	𝍸 𝍸 𝍸 |||	18
علي	𝍸 𝍸 𝍸 𝍸	20
المجموع		85

يلاحظ أنه من خلال الجداول السابقة، يمكن تلخيص البيانات وتنظيمها وتبويبها بحيث تبدو ذات معنى، إذ يمكن بيان الحالات وعدد التكرارات التي تملكها كل حالة ونسبتها المئوية بالنسبة لمجموع الحالات الكلية، الأمر الذي يساعد بالتالي في ترتيبها تنازلياً أو تصاعدياً في ضوء عدد التكرارات المقابلة لها، كما تتيح مثل هذه الجداول إمكانية إجراء بعض العمليات الاحصائية على البيانات والتمثيل البياني لها.

2- الجداول التكرارية التلخيصية

تعد هذه الجداول الأكثر استخداماً وشيوعاً في عمليات تلخيص البيانات الاحصائية وتبويبها، ولا سيما في حالة كون البيانات الاحصائية ترتبط بعدد من المتغيرات ذات علاقة بأفراد المجتمع الاحصائي موضع البحث، أو عندما تتوفر لدينا قاعدة بيانات كبيرة حول متغير أو عدد من المتغيرات الاحصائية. ففي بعض الحالات يتم جمع بيانات عن أفراد مجتمع ما تتعلق بأكثر من خاصية أو متغير واحد، الأمر الذي يجعل من استخدام الجداول التكرارية البسيطة لتلخيص مثل هذه البيانات عملية غير مفيدة. ففي

مثل هذه الحالة يقتضي الأمر تنظيم البيانات وتصنيفها في ضوء متغيرات البحث ومستوياتها المختلفة وذلك حتى تُسهل عملية معالجتها وإجراء العمليات الاحصائية المطلوبة عليها. ولتحقيق هذا الغرض يمكن استخدام الجداول التكرارية التلخيصية، ومثل هذه الجداول تشتمل علي عدد من الخلايا تمكن من عرض المستويات المختلفة من المتغيرات وعدد التكرارات المقابلة لها بالإضافة إلى النسب المئوية والمجموع الكلي لكل متغير على حدة.

ولتوضيح مثل هذه الجداول، يمكن ملاحظة المثال (2:4) والذي يعطي بياناً تفصيلاً لعدد الطلبة المتقدمين لامتحان الشهادة الثانوية في الأردن وفقاً لمتغير الجنس والمنطقة والفرع الدراسي.

جدول (2:4): عدد المتقدمين لامتحان الشهادة الثانوية في الأردن حسب المنطقة الجغرافية والجنس والفرع الدراسي.

عدد المتقدمين / المنطقة	الفرع العلمي			الفرع الأدبي			الفرع المهني		
	ذكور	إناث	المجموع	ذكور	إناث	المجموع	ذكور	إناث	المجموع
عمان	8200	7800	16000	7900	7200	15100	2100	3000	5100
اربد	6500	6800	13300	5400	5800	11200	1800	1900	3700
الكرك	1100	0900	02000	0980	01100	2080	230	350	5800
عجلون	0600	0650	01150	2240	1380	3620	240	320	5600
المجموع	16400	16150	32450	16520	15480	32000	4370	5570	20200

لاحظ أن مثل هذا النوع من الجداول يمكننا من التعامل مع عدد كبير من البيانات التي ترتبط بأكثر من متغير واحد بالوقت نفسه. ففي مثالنا السابق لاحظ أن البيانات التي تم جمعها حول المتقدمين لامتحان الشهادة الثانوية تم تصنيفها وتلخيصها وفق ثلاثة متغيرات ولكل منها عدد من المستويات وهو متغير المنطقة الجغرافية وتشمل محافظات المملكة، ومتغير الجنس ويتضمن الذكور والإناث ثم متغير الفرع الأكاديمي وله ثلاثة مستويات هي: الفرع العلمي، والأدبي، والمهني.

بالإضافة لما سبق يمكن استخدام الجداول التلخيصية للتعامل مع البيانات الكبيرة التي ترتبط بمتغير كمي معين، فمن غير المجدي التعامل مع البيانات على نحو منفرد لأن ذلك سيؤدي إلى الحصول على قائمة طويلة من البيانات يصعب التعامل معها ومعالجتها. وعوضاً عن ذلك يتم تلخيص هذه البيانات وتبويبها وفق عدد من الفئات اعتماداً على طول فئة محدد يتم حسابه لهذا الغرض. وفي اعداد مثل هذه الجداول يتم اتباع الخطوات التالية:

1- تحديد أدنى قيمة وأعلى قيمة في التوزيع .

2- تحديد عدد الفئات المطلوبة ، بحيث يجب أن يكون عددها معقولاً . فالعدد القليل من الفئات ربما يؤدي إلى إضاعة بعض البيانات ، أما العدد الكبير لها لا يؤدي إلى تلخيصها . ويرى معظم الاحصائيين أن العدد المفضل لعدد الفئات ذلك الذي يتراوح بين 10و 15 فئة (النبهان ، 2001) .

3- تحديد المدى وذلك من خلال ايجاد الفرق بين أعلى قيمة وأدنى قيمة في التوزيع .

4- تحديد طول الفئة وذلك بقسمة المدى على عدد الفئات المطلوبة . هذا ويفضل أن يكون طول الفئة عدداً فردياً من أجل أن يكون مركز الفئة عدداً صحيحاً . كما يمكن تقريب طول الفئة ليصبح عدداً صحيحاً وذلك عندما يكون الناتج كسراً .

5- تحديد حدود الفئات . تعرف حدود الفئات بالحد الذي تبدأ به الفئة (الحد الأدنى) والحد الذي تنتهي به (الحد الأعلى) . يمكن للباحث اعتبار أقل قيمة في التوزيع أو قيمة أدنى منها لا توجد في التوزيع على أنها الحد الادنى للفئة ، ولكن يفضل أن يكون الحد الأدنى للفئة من مضاعفات طول الفئة ، شريطة أن يكون أقل من أدنى قيمة في التوزيع . أما الحد الأعلى للفئة فهو بمثابة الحد الأدنى للفئة مضافاً إليه طول الفئة ، مطروحاً من ذلك وحدة القياس . فعلى سبيل المثال ، إذا كان طول الفئة (5) وكانت أقل مشاهدة في التوزيع (3) فيمكن تشكيل الفئات على النحو الآتي :

الحد الأدنى		الحد الأعلى		الحد الأدنى		طول الفئة		وحدة القياس	
2	-	⑥	⟵	2	+	5	-	1 =	⑥
أو									
3	-	⑦	⟵	3	+	5	-	1 =	⑦

لاحظ أنه في الحالة الأولى كان الحد الأدنى للفئة (2) وهو أقل من أدنى قيمة في التوزيع وهي(3) ، أما في الحالة الثانية فكان الحد الأدنى للفئة (3) وهي أقل قيمة في التوزيع . وتجدر الإشارة هنا أن كلا الحدين الأدنى والأعلى هما ينتميان للفئة . وفي مثل هذا المثال ، لا يمكن اعتبار أحد مضاعفات طول الفئة على أنه الحد الأدنى للفئة لأن هناك قيماً في التوزيع أقل من مضاعفات طول الفئة . فاعتبار أحد مضاعفات طول الفئة كحد أدنى للفئة في مثالنا هذا يؤدي إلى ضياع بعض المشاهدات وذلك كما هو موضح أدناه .

❖ الفئة 10 - 14 : العدد (10) هو أحد مضاعفات طول الفئة (5)، فاعتباره حد أدنى للفئة يؤدي إلى ضياع قيم في التوزيع مثل (3) وغيرها من القيم الأخرى التي تقل عن (10) .

مثال (2:5) : لنفرض أن لدينا توزيع من الدرجات يتراوح عددها (30) درجة . فإذا كانت أقل قيمة في هذا التوزيع (أدنى درجة) هي (12) وكان طول الفئة المطلوب (5)، فما هو الحد الأدنى والأعلى للفئة الأولى في التوزيع ؟

الحل: في حالة هذا المثال يمكن اعتبار أحد مضاعفات طول الفئة كحد أدنى لهذه الفئة على اعتبار عدم وجود قيم (درجات) أقل منه في التوزيع وذلك على النحو الآتي :

الحد الأدنى للفئة $(2 \times 5) = (10)$

الحد الأعلى للفئة = الحد الأدنى + طول الفئة – وحدة القياس

$= 10 + 5 - 1 = 14$

وعليه تكون الفئة الأولى لهذا التوزيع على النحو الآتي :

(10 - 14) لاحظ أن هذه الفئة تشتمل على أقل قيمة في التوزيع وهي (12) .

يمكن أيضاً أن تبدأ هذه الفئة بأقل قيمة في التوزيع كحد أدنى لها وذلك على النحو الآتي :

الحد الأدنى (12) : وهي أقل قيمة في التوزيع

الحد الأعلى $= 12 + 5 - 1 =$ (16)

وعليه تكون الفئة الأولى للتوزيع هي(12-16) .

6- تحديد عدد التكرارات المقابلة لكل فئة من الفئات وذلك بحصر عدد المشاهدات في التوزيع التي تنتمي لكل فئة .

مثال (2:6) : طبق باحث مقياساً للقلق على عينة من الأفراد بلغ حجمها (50) فرداً وكانت درجاتهم على هذا الاختبار كما هو في التوزيع الآتي :

62, 61, 43, 40, 21, 30, 29, 17, 16, 12, 11, 11

34, 32, 12, 23, 22, 28, 51, 52, 56, 14, 13, 15, 15

12, 11, 16, 18, 27, 20, 19, 17, 14, 50, 54, 38, 43

58, 48, 53, 21, 25, 32, 24, 33, 36,41, 59, 60

المطلوب : بناء جدول تكراري لهذه البيانات؟

الحل :

(1) نحدد أدنى قيمة وأعلى قيمة في التوزيع وهما على الترتيب 11 و 62 .

(2) نجد الفرق بين أعلى قيمة وأدنى قيمة ونحسب طول الفئة وذلك بقسمة هذا الفرق على عدد الفئات المنوي تكوينها .

أعلى قيمة = 62 ، أدنى قيمة = 11

الفرق = 62 - 11 = 51

عدد الفئات المفترض = 10

∴ طول الفئة = $\frac{51}{10}$ = 5.1 نقرب هذا العدد لأقرب عدد صحيح وهو (5) . وبذلك يصبح طول الفئة المطلوب هو (5) .

(3) نحدد الحد الأدنى والحد الأعلى للفئة الأولى

الحد الأدنى (10) وهو أحد مضاعفات طول الفئة

الحد الأعلى = الحد الأدنى + طول الفئة – وحدة القياس

= 10 + 5 - 1 = 14

(4) بناء جدول يتضمن عمودين ، أحدهما يمثل الفئات والأخر يمثل التكرارات المقابلة لها ، بحيث يتم رصد عدد الحالات التي تقع مقابل كل فئة في التوزيع ، وذلك من خلال قراءة كل مشاهدة في التوزيع ووضع إشارة (/) مقابل الفئة التي تنتمي لها . وفي حال الانتهاء من رصيد جميع المشاهدات ، يتم جمعها لايجاد التكرارات المقابلة لكل فئة ، وذلك كما هو مبين في الجدول التالي :

الفئات	الاشارات	التكرارات (ت)
14 - 10	𝍸 ||||	9
19 - 15	𝍸 |||	8
24 - 20	𝍸 |	6
29 - 25	||||	4
34 - 30	𝍸	5
39 - 35	||	2
44 - 40	||||	4
49 - 45	|	1
54 - 50	𝍸	5
59 - 55	|||	3
64 - 60	|||	3
المجموع		50

يمكن لمثل هذه الجداول أن تشتمل أعمدة إضافية أخرى لعرض معلومات ذات علاقة مثل النسب المئوية ومركز الفئة والتكرار التراكمي ، أو أية معلومات أخرى من شأنها أن تساعد في إجراء العمليات الاحصائية المطلوبة وذلك تبعاً للأهداف التي من أجلها جمعت مثل هذه البيانات .

التكرار النسبي للفئة Interval Relative Frequency:

يمثل التكرار النسبي للفئة عدد الحالات (المشاهدات) التي تشتمل عليها الفئة مقسوماً على المجموع الكلي للتكرارات في التوزيع مضروباً في 100% . وفي مثالنا السابق نلاحظ أن التكرار النسبي للفئة الخامسة (30 - 34) هو $\frac{5}{50}$ ويساوي 10% .

التكرار التراكمي للفئة Cumulative Frequency:

يمثل التكرار التراكمي للفئة مجموع التكرارات المقابلة لها مضافاً إليه مجموع التكرارات التراكمية للفئة التي تسبقها . ويرمز للتكرار التراكمي بالرمز (ت ت) ويتطلب حسابه ترتيب البيانات الاحصائية تنازلياً أو تصاعدياً . ويفيد التكرار التراكمي في حساب المئينات أو النسب المئوية للحالات التي تقابل فئة

معينة أو تقع دونها، كما أنه يساعدنا في حساب الوسيط والرتب الميئنية للمشاهدات في توزيع احصائي معين. وفي حالة المثال السابق يمكن تحديد التكرار التراكمي للفئات على النحو الآتي :

الفئة	التكرار	التكرار التراكمي
14 - 10	9	9
19 - 15	8	(9 + 8) : 17
24 - 20	6	(6 + 17) : 23
29 - 25	4	(4 + 23) : 27
34 - 30	5	(5 + 27) : 32
39 - 35	2	(2 + 32) : 34
44 - 40	4	(4 + 34) : 38
49 - 45	1	(1 + 38) : 39
54 - 50	5	(5 + 39) : 44
59 - 55	3	(3 + 44) : 47
64 - 60	3	(3 + 47) : 50
المجموع	50	

مركز الفئة Mid of Interval:

يعرَّف مركز الفئة على أنه متوسط الفئة ، وهو بمثابة مجموع كل من الحد الأدنى والحد الأعلى للفئة مقسوماً على (2)، ويرمز له بالرمز (م س).

فمثلاً لو أخذنا الفئة التالية (30 - 34)، نجد أن مركز الفئة لها (م س) هو 30 + 34 ÷ 2 = (32)
وبالرجوع إلى المثال (2:5) يمكن توضيح مراكز الفئات على النحو المبين في الجدول التالي :

*لاحظ أعلاه، أن التكرار التراكمي لكل فئة هو عبارة عن عدد التكرارات المقابلة لها بالإضافة إلى التكرارات التراكمية المقابلة للفئة التي تسبقها.

الفئة	مركز الفئة (م س*)	التكرار (ت)
14 - 10	12	9
19 - 15	17	8
24 - 20	22	6
29 - 25	27	4
34 - 30	32	5
39 - 35	37	2
44 - 40	42	4
49 - 45	47	1
54 - 50	52	5
59 - 55	57	3
64 -60	62	3
المجموع		50

الحدود الفعلية للفئات Real Interval limits

لكل فئة من فئات التوزيع هناك حدان فعليان؛ احدهما يعرف بالحد الفعلي الأدنىLower real limit والآخر يسمى بالحد الفعلي الأعلىupper real limit. ويمثل الحد الفعلي الأدنى للفئة حدها الأدنى مطروحاً منه نصف وحدة القياس، في حين يمثل الحد الفعلي لها الحد الأعلى مضافاً إليه نصف وحدة القياس، وذلك كما هو مبين في الجدول التالي:

*لاحظ أن الفرق بين مركز الفئة ومركز الفئة التي يسبقها يساوي دائما طول الفئة، فهو (5) في مثالنا هذا.

الحدود الفعلية	الفئات
9.5 - 4.5	9 - 5
14.5 - 9.5	14 - 10
19.5 - 14.5	19 - 15
24.5 - 19.5	24 - 20
19.5 - 24.5	29 - 25

لاحظ في الجـدول أعلاه أنه تم طرح نصـــف وحـــدة القيـاس من حـــدود الفـئات الدنيـا وهـــي (5 ,10 ,15 ,20 ,25) وذلك لتـحـــديد الحـــدود الفـعليـــة الدنيـا لهـا بحـيث أصـبـــحت (4.5 ,9.5 ,14.5 ,19.5 ,24.5) . أما في حالة تحديد الحدود الفعلية العليا لتلك الفئات ، فقد تم إضـافة نصف وحدة القيـاس لحدودها العليا (9 ,14 ,19 ,24 ,29)، لتـصبح على النحـو الآتي (9.5 ,14.5 ,19.5 ,24. 5 ,29.5).

وتجدر الإشارة هنا ، أن كتابة الفئات بدلالة حدودها الفعلية لا يؤثر في طول الفئة لهذه الفئات ، إذ أنها تحتفظ بنفس طول الفئة ، ويمثل حدها الفعلي الأعلى مطروحاً منه حدها الفعلي الأدنى . فلو نظرنا إلى الفئة التالية على سبيل المثال (14.5 - 19.5)، نجد أن طول الفئة هو عبارة عـن ناتج الفـرق بين (14.5) و (19.5) وهو مساوياً لـ (5) وحدات .

تستخدم الحدود الفعلية عادة في حالة المتغيرات المستمرة أو المتصلة وذلك لأن مثل هذه المتغيرات تشتمل عدداً كبيراً من القيم في مدى معين ، وقد تأخذ قيماً كسرية في ذلك المدى ، وهذا يجعل من عملية قياسها أمراً غير دقيق . ومن أجل تخطي اخطاء القياس وتحقيق مزيد من الدقة على نحو يجعل الفئات تشتمل على جميع القيم المحتملة للمتغير ، فبات من الضروري استخدام فكرة الحدود الفعلية في حال قياس مثل تلك المتغيرات .

ثانياً : العرض البياني للبيانات

بالإضافة إلى استخدام الجداول لعرض البيانات الاحصائية ، يمكن استخدام العرض البياني لتمثيل تلك البيـانات . فـالعـرض البـيـاني يعطينا صـورة أوليـة عن شكل توزيع البـيـانات ومـدى تشـتـتـهـا أو تجمعها . وهذا بالطبع يساعد في وصف خصائص المجتمع وربما عمل بعض الاستنتاجات والأحكام حول تلك الخصائص .

تختلف أساليب العرض البياني باختلاف طبيعة المتغيرات موضع البحث من حيث كونها كمية (منفصلة، مستمرة) أو نوعية، بالإضافة إلى حجم البيانات التي يتم جمعها عن تلك المتغيرات. وفيما يلي عرض لأساليب العرض البياني.

1 - القطع الدائري Pie Chart:

يستخدم القطع الدائري لتمثيل البيانات المتعلقة بمتغيرات نوعية لبيان المستويات أو الأوجه المختلفة التي تتخذها. وتتطلب هذه الطريقة تقسيم الدائرة إلى قطاعات مختلفة تتناسب مع التكرارات المناظرة لها والتي تقابل كل قطاع من هذه القطاعات. ولتوضيح ذلك، يمكن أخذ المثال التالي:

مثال (2:6): يلخص الجدول التالي اعداد الطلبة المتقدمين لامتحان الشهادة الثانوية في خمس محافظات في الأردن مقرباً لأقرب ألف. المطلوب هو تمثيل ذلك من خلال القطع الدائري:

المحافظة	عدد المتقدمين
عمان	12.000
اربد	10.000
الزرقاء	8.000
الكرك	6.000
جرش	4.000
المجموع	40000

الحل: لتمثيل هذه البيانات من خلال القطع الدائري، يمكن اتباع الخطوات التالية:

1- حساب النسبة المئوية (التكرار النسبي) لكل وجه من أوجه المتغير وذلك من خلال قسمة عدد المتقدمين على المجموع الكلي مضروباً في 100% وذلك على النحو الآتي:

عمان : $\frac{12.000}{40.000} \times 100\% = 30\%$

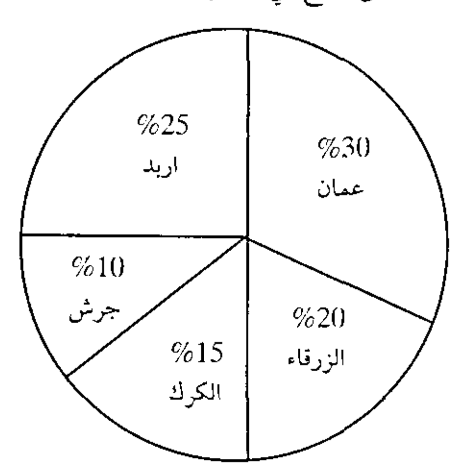
اربد : $\frac{10.000}{40.000} \times 100\% = 25\%$

الزرقاء : $\frac{8.000}{40.000} \times 100\% = 20\%$

الكرك : $\frac{6.000}{40.000} \times \%100 = \%15$

جرش : $\frac{4000}{40.000} \times \%100 = \%10$

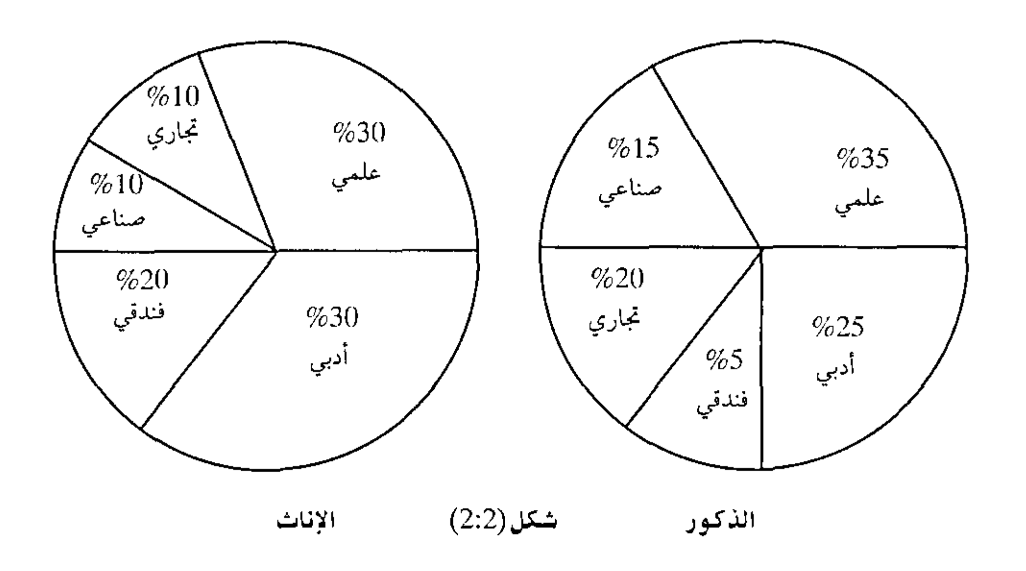
2- تحديد الزاوية المقابلة لنسبة كل قطاع في الدائرة وذلك على النحو الآتي:



شكل (2:1) يوضح هذا التمثيل البياني المستويات المختلفة لمتغير عدد الطلبة المتقدمين لامتحان الشهادة الثانوية

يمكن استخدام القطع الدائري أيضاً للمقارنة بين متغيرين نوعيين في ضوء عدد من المستويات وذلك من خلال تمثيل المستويات المختلفة لكل متغير في قطع دائري منفصل. وذلك كما هو موضح في المثال التالي:

مثال (2:7): الشكل التالي يوضح اعداد الطلبة الملتحقين في الفروع الدراسية المختلفة حسب متغير الجنس.



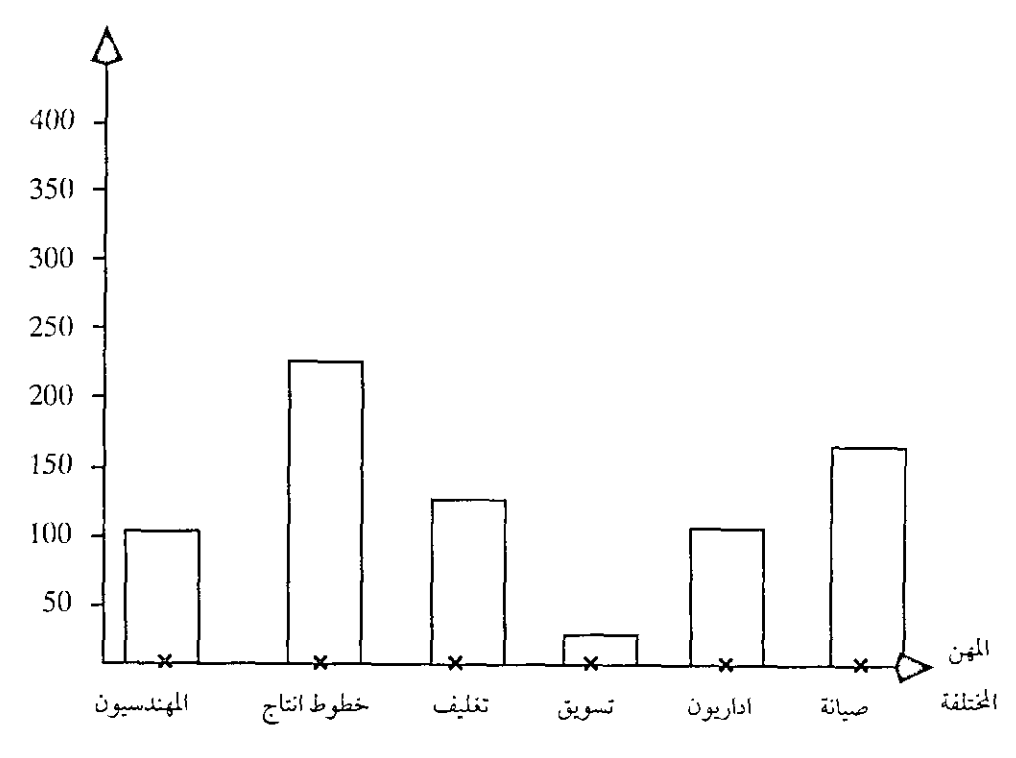
الإناث شكل (2:2) الذكور

2- **التمثيل بالأعمدة Bar Graph:**

كما هو الحال في التمثيل من خلال القطع الدائري ، يمكن استخدام الأعمدة لبيان التوزيع التكراري أو التوزيع النسبي التكراري لمجموعة أو أكثر من المتغيرات أو المستويات المتعددة للمتغير الواحد. وتستخدم عادة الأعمدة في بيان التوزيعات التكرارية للمتغيرات النوعية وكذلك الكمية منها ولا سيما المتغيرات الكمية المنفصلة.

يتطلب التمثيل البياني للبيانات من خلال الأعمدة – رسم خطين متعامدين بزاوية مقدارها (90°)، بحيث يمثل الخط الأفقي المتغيرات المختلفة أو المستويات المتعددة من متغير نوعي ما ، في حين يمثل الخط العمودي التوزيعات التكرارية أو التوزيعات التكرارية النسبية المقابلة لها ، وذلك كما هو موضح بالأمثلة التالية :

مثال (2:8) : عدد العاملين في المهن المختلفة في قطاع إنتاجي معين مقرباً لأقرب مئة.



شكل (2:3) : التمثيل بالأعمدة للبيانات الإحصائية

❖ يمكن تحويل التوزيعات التكرارية إلى توزيعات تكرارية نسبية في حال كون حجم البيانات كبيراً وذلك من اجل تسهيل عملية تمثيلها من خلال الأعمدة ، وذلك كما هو موضح في المثال التالي :

مثال (2:9) : يظهر الجدول التالي أعداد السكان مقرباً لأقرب مليون في دول بلاد الشام . والمطلوب تمثيل ذلك بيانياً من خلال الأعمدة مستخدماً التوزيعات التكرارية النسبية لعدد السكان في كل دولة .

الدولة	عدد السكان
سوريا	22.000000
الأردن	6.000000
فلسطين	4.000000
لبنان	8.000000
المجموع	40.000000

لعرض هذه البيانات من خلال الأعمدة باستخدام التوزيعات التكرارية النسبية يتطلب ذلك إجراء الخطوات التالية :

1- تحويل التوزيعات التكرارية المقابلة لكل دولة إلى توزيعات تكرارية نسبية من خلال قسمة عدد السكان في كل دولة على المجموع الكلي لعدد السكان وضرب الناتج في 100% وذلك على النحو الآتي :

التوزيع التكراري النسبي لعدد السكان في سوريا = $\frac{22.000000}{40.000000} \times 100\% = 55\%$

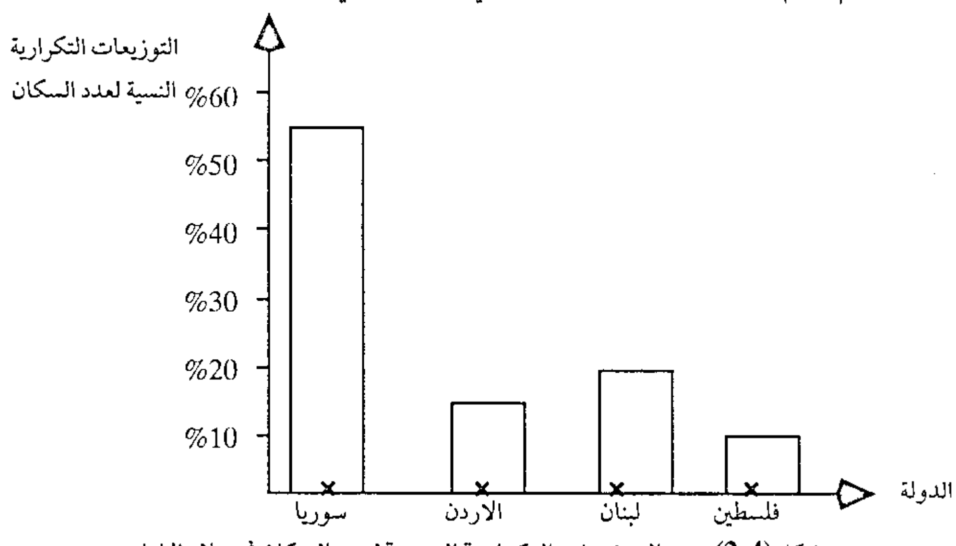
التوزيع التكراري النسبي لعدد السكان في الأردن = $\frac{6.000000}{40.000000} \times 100\% = 15\%$

التوزيع التكراري النسبي لعدد السكان في فلسطين = $\frac{4.000000}{40.000000} \times 100\% = 10\%$

التوزيع التكراري النسبي لعدد السكان في لبنان = $\frac{8.000000}{40.000000} \times 100\% = 20\%$

2-رسم خطين متعامدين بزاوية مقدارها 90°، بحيث يمثل الخط الأفقي المتغير بمستوياته المتعددة (عدد السكان في كل من سوريا، الأردن، لبنان، وفلسطين)، في حين يمثل الخط العمودي التوزيعات التكرارية النسبية لعدد السكان في كل دولة .

3-من خلال الإشارة بنقطة على الشكل يتم تحديد التوزيع النسبي المقابل لعدد السكان في كل دولة، ومن ثم رسم الأعمدة وذلك كما هو مبين في الشكل التالي :



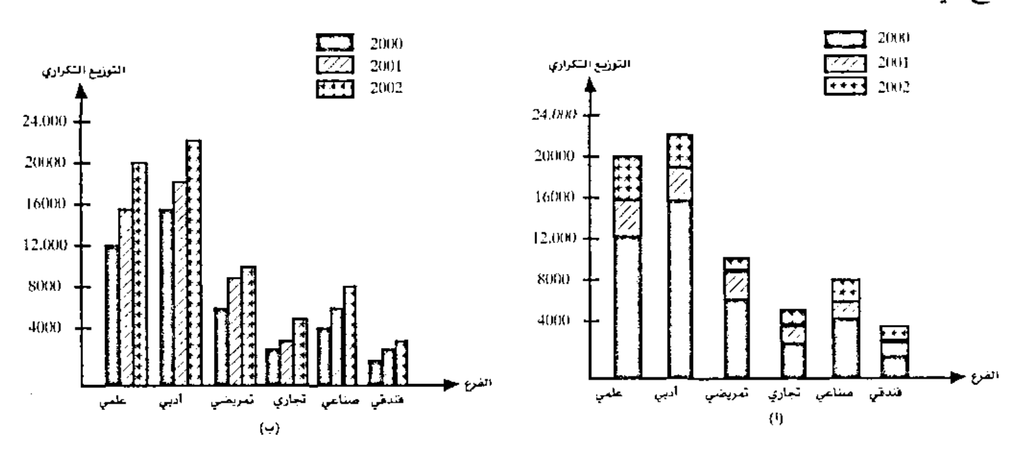
شكل (2:4) يبين التوزيعات التكرارية النسبية لعدد السكان في بلاد الشام

إن التمثيل من خلال الأعمدة يبرز الفروق بين المستويات المختلفة لمتغير ما كما هو موضح في الشكل أعلاه، ويمكن أيضاً أن يبين التغيرات التي تطرأ على المتغير الواحد مقارنة بالمتغيرات الأخرى وذلك من خلال استخدام نفس العمود الواحد وتجزئته إلى عدة أجزاء او من خلال استخدام الأعمدة المتلاصقة كإشارة للتغير أو التطور الذي حدث عليه. ويمكن توضيح ذلك من خلال المثال التالي:

مثال (2:10) : يبين الجدول التالي أعداد الطلبة من مختلف التخصصات الذين تقدموا لامتحان الشهادة الثانوية خلال الأعوام 2000, 2001, 2002 والمطلوب تمثيل ذلك بيانياً من خلال الأعمدة.

الفرع	أعداد الطلبة خلال السنوات			المجموع
	2000	2001	2002	
العلمي	12000	15000	20000	47000
الأدبي	15000	18000	22000	55000
التمريضي	6000	9000	10000	25000
التجاري	2000	3000	4200	9200
الصناعي	4000	6000	7000	17000
الفندقي	1500	1800	2300	5600
المجموع	40.500	52.800	65.500	158.800

يمكن تمثيل هذه البيانات من خلال استخدام الأعمدة المجزأة أو الأعمدة المتلاصقة وذلك كما هو موضح في الشكل (2:5 أ ، ب)



شكل (2:5) يوضح التوزيعات التكرارية لعدد الطلاب المتقدمين لامتحان الشهادة الثانوية من مختلف الفروع خلال الأعوام الدراسية 2000, 2001, 2002 وذلك من خلال التمثيل بالأعمدة المجزأة (أ) والأعمدة المتلاصقة (ب)

3-المدرج التكراري Histogram

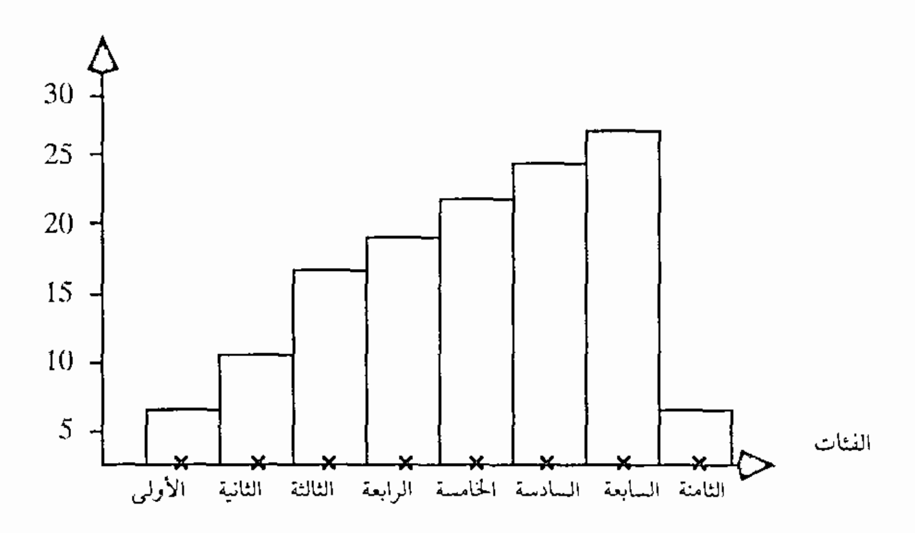
يستخدم المدرج التكراري لعرض البيانات المتعلقة بالمتغيرات الكمية ولا سيما المستمرة منها، وهو عبارة عن مجموعة من مستطيلات متلاصقة تمثل قاعدتها الفئات في حين يمثل ارتفاعها تكرار هذه الفئات. ولتوضيح ذلك يمكن الاستعانة بالمثال التالي:

مثال (2:11): يظهر الجدول التالي معاملات الذكاء لعينة من الأفراد عددها 100 طالب. المطلوب تمثيلها بيانياً من خلال المدرج التكراري.

جدول يوضح معاملات الذكاء لعينة من الأفراد عددها (100)

الفئات	التكرارات
59 - 50	4
69 - 60	7
79 - 70	10
89 - 80	14
99 - 90	17
109 - 100	21
119 - 109	23
129 - 120	4
المجموع	100

لتمثيل هذه البيانات من خلال استخدام المدرج التكراري، يتطلب الأمر رسم محورين متعامدين، يمثل المحور الأفقي متغير الذكاء (فئات الذكاء)، في حين يمثل المحور العمودي التكرارات المقابلة للفئات وذلك على النحو التالي:

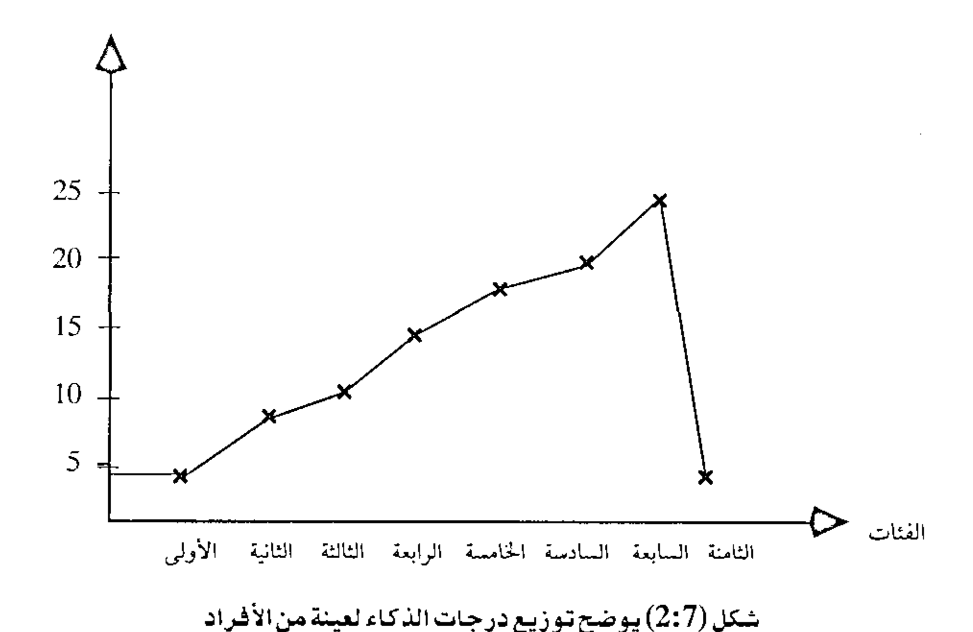


شكل (2:6) توزيع درجات ذكاء عينة من الأفراد عددها (100)

وتجدر الإشارة هنا، إن أطوال قاعدة الأعمدة التي تمثل الفئات يجب أن تكون متساوية على اعتبار تساوي طول الفئة لمثل هذه الفئات، هذا ويمكن استخدام الحدود الفعلية أو مراكز الفئات في المحور الأفقي للدلالة على الفئات، بحيث يكون مركز الفئة في منتصف قاعدة العمود الدال عليها.

4- **المضلع التكراري** Frequency Polygon

يماثل المضلع التكراري المدرج التكراري من حيث أنه يستخدم للتعبير عن المتغيرات الكمية، ويكمن الفرق بينهما من حيث أنه في المضلع التكراري يتم الاستعاضة عن الأعمدة بوضع نقطة مقابل تكرار كل فئة ويتم ايصال هذه النقاط معاً بخطوط مستقيمة، وذلك كما يتضح في الشكل التالي وهو بمثابة تمثيل بياني لنسب معاملات الذكاء الواردة في المثال رقم (2:11):



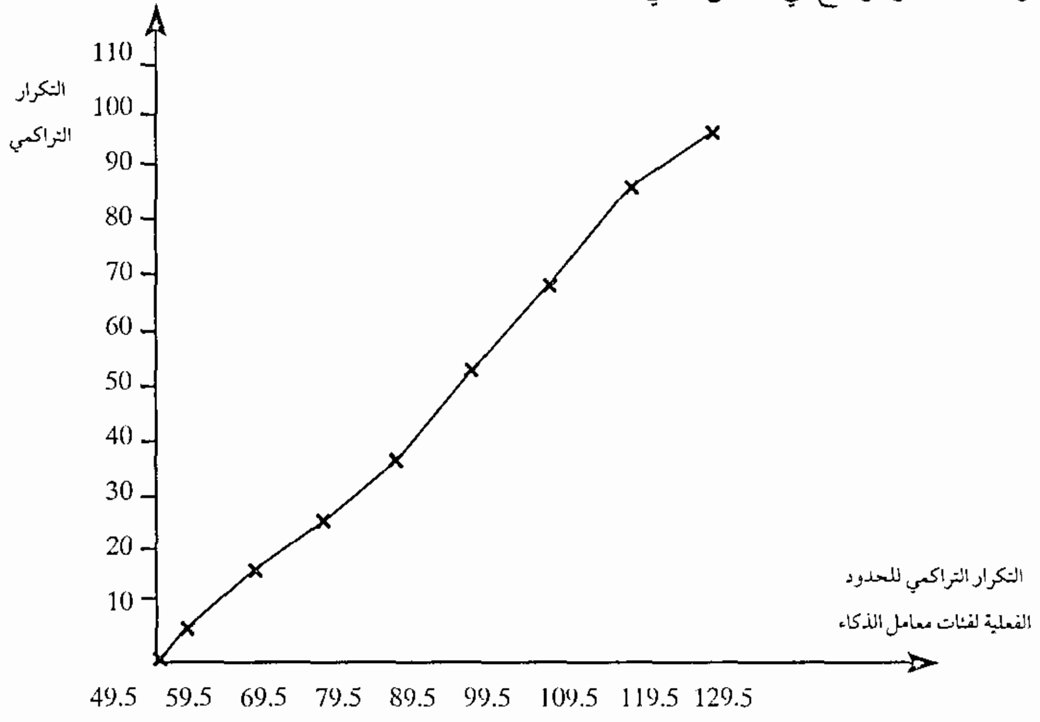
شكل (2:7) يوضح توزيع درجات الذكاء لعينة من الأفراد

5- **المضلع التكراري التراكمي** Cumulative Frequency Polygon:

هذا النوع من التمثيل البياني للبيانات يشبه إلى درجة كبيرة المضلع التكراري، ولكن يتطلب تعيين التكرار التراكمي لكل فئة من خلال تحديد عدد الحالات (التكرارات) التي تشتمل عليها الفئة وتلك التي تقابل الفئات التي تسبقها. ولتوضيح ذلك يمكن الرجوع إلى البيانات الواردة في المثال (2:11).

الفئات	التكرار	التكرار التراكمي
59 - 50	4	4
69 - 60	7	11
79 - 70	10	21
89 - 80	14	35
99 - 90	17	52
109 - 100	21	73
119 - 109	23	96
129 - 120	4	100
المجموع	100	

ولتمثيل تلك البيانات باستخدام المضلع التكراري التراكمي يستلزم الأمر رسم محورين متعامدين، يمثل المحور الأفقي الحدود الفعلية للفئات بدءاً بالحد الفعلي الأدنى لأول فئة في التوزيع وانتهاء بالحد الفعلي الأعلى للفئة الأخيرة، في حين يمثل المحور العمودي التكرارات التراكمية (المتجمعة) لهذه الفئات وذلك كما هو موضح في الشكل التالي:



شكل (2:8) المضلع التكراري التراكمي

6- الأغصان والأوراق Stem & Leaf

تعد طريقة الغصن والأوراق إحدى الطرق التي تستخدم لتمثيل البيانات المتعلقة بالمتغيرات الكمية وهي شأنها شأن طرق المدرج والمضلع التكراري في كونها تبرز شكل التوزيع ومدى تجمع أو تشتت المشاهدات والتغير الذي يطرأ عليها، لكنها تعد طريقة سريعة وسهلة لا تتطلب مهارات الرسم أو الحساب. تصلح هذه الطريقة في حالة المشاهدات التي لا يزيد عددها عن بضع مئات والتي تتألف من منزلتين أو أكثر.

ولتمثيل البيانات من خلال هذه الطريقة يتطلب الأمر رسم خط عمودي بحيث يتم رصد المنازل العشرية أو المئوية للأعداد على يساره لتمثل الغصن، في حين يتم رصد المنازل الأحادية على يمينه لتمثل الأوراق ، وذلك على النحو المبين في المثال التالي :

مثال (2:12) : أمامك توزيع درجات (100) طالب على امتحان مستوى اللغة الانجليزية على النحو الآتي :

16, 16, 16, 15, 14, 13, 13, 11, 10, 10, 09, 07 21, 21, 21, 21,
21, 20, 20, 20, 19, 19, 18, 34,17, 29, 29, 27, 26, 25, 25, 24,
24, 24, 23, 22, 22, 36, 35, 35, 35, 34, 33, 32, 31, 30, 30, 30,
44, 43, 42, 42, 41, 40, 40, 40, 39, 38, 38, 37, 54, 53, 53, 53,
52, 51, 51, 51, 49, 48, 46, 45, 67, 66, 66, 64, 64, 63, 61, 60,
59, 58, 58, 56, 83, 81, 79, 77, 77, 75, 73, 72, 70, 70, 69, 68,
95, 94, 92, 87.

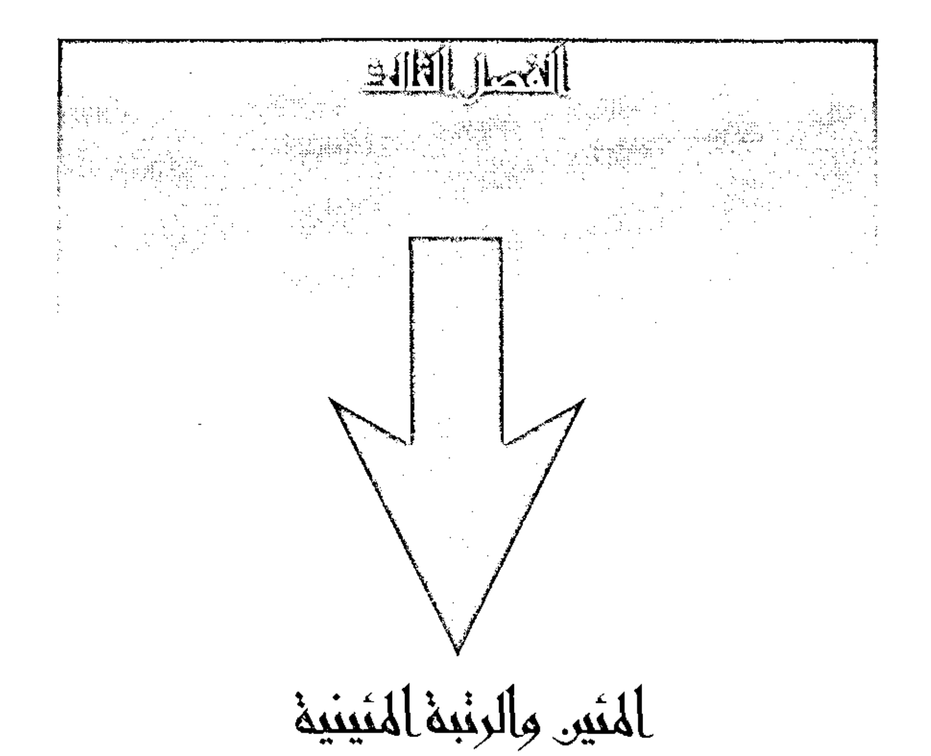
لاحظ أن أدنى درجة في التوزيع هي (7) في حين أعلى درجة كانت 95. وبذلك الناتج يكون لدينا (10) أغصان بدءاً من الصفر وانتهاءً بالعدد (9) وذلك كما هو مبين في الشكل التالي :

الساق و(الغصن)	0	7 9	
	1	0 0 1 3 3 4 5 6 6 6 7 8 9 9	
	2	0 0 0 1 1 1 1 1 2 2 3 4 4 4 5 5 6 7 9 9	
	3	0 0 0 1 2 3 4 4 5 5 5 6 7 8 8 9	
	4	0 0 0 1 2 2 3 4 5 6 7 9	
	5	1 1 1 2 3 3 3 4 6 8 8 9	الأوراق
	6	0 1 3 4 4 6 6 7 8 9	
	7	0 0 2 3 5 7 7 9	
	8	1 3 7	
	9	2 4 5	

شكل (2:9)

وبالنظر إلى التوزيع أعلاه نلاحظ أن عدداً كبيراً من درجات الطلبة كانت دون الـ (40)، وأن عدداً قليلاً منها كان في مستوى الثمانون أو أكثر. ولتسهيل قراءة مثل هذا التمثيل، فلنأخذ المثال التالي:

لاحظ أنه في توزيع درجات الطلاب إن هناك خمسة طلاب حصلوا على علامة (21)، وبالنظر إلى التمثيل البياني (الغصن والأوراق) نلاحظ إن هناك خمس واحدات تقع على يمين الرقم (2)، وهذا يعني أن درجة (21) تكررت خمس مرات، وهكذا يمكن قراءة بقية الدرجات على هذا النحو.



المئين والرتبة المئينية

Percentile and Percentile Ranks

تمهيد:

قد يتطلب الأمر أحياناً الحكم على قيمة معينة في توزيع ما من حيث تحديد موقعها النسبي بالنسبة لمجموعة القيم التي يشتمل عليها ذلك التوزيع ، فعلى سبيل المثال ، قد يحصل طالب على درجة في امتحان ما مقدارها (60) ، فمثل هذه الدرجة لا يمكن الحكم عليها ما إذا كانت مرتفعة أو منخفضة بمجرد النظر إليها ، وإنما يقتضي الأمر مقارنتها مع مجموعة الدرجات الأخرى التي حصل عليها الطلاب الآخرون الذين تقدموا لذلك الامتحان . ومن أجل تحديد الموقع النسبي لهذه الدرجة يتم اللجوء عادة إلى حساب المئين والرتبة المئينية .

يعرّف المئين (ي) بأنه الدرجة التي تقع دونها أو أعلى منها نسبة معينة من الدرجات في توزيع ما . فالمئين هو بمثابة علامة في توزيع معين يقل عنها نسبة محددة من الحالات (الدرجات) بالوقت الذي يوجد نسبة أعلى منها من الدرجات في ذلك التوزيع . فعلى سبيل المثال ، إذا كانت علامة طالب في امتحان ما هي (70) وكان المئين الموازي لها هو (60) ، فهذا يعني أن 60% من الدرجات تقع دون هذه العلامة ، وأن 40 % من الدرجات في ذلك التوزيع تقع أعلى منها .

تستخدم المئينات بشكل واسع لتحديد المواقع النسبية للحالات الاحصائية وفي المقارنات بين أداء الأفراد في العديد من اختبارات التحصيل والذكاء والشخصية بمختلف أنواعها وذلك لأنها توفر معايير موحدة للمقارنة تتمثل بالنسب المئوية لعدد الحالات التي تقع دون أية حالة من الحالات ، ولتوضيح ذلك فلنأخذ المثال التالي : إذا كان المئين (75) يساوي (45) ، فهذا يعني أن 75 % من عدد الحالات تقع دون الدرجة (45) ، في حين 25 % من عدد الحالات تلك تقع فوق هذه الدرجة .

حساب المئينات:

عموماً يتطلب تحديد المئين ترتيب البيانات الاحصائية تنازلياً أو تصاعدياً وتعيين التكرارات التراكمية المقابلة لهذه البيانات بحيث يصار إلى تحديد موقع التكرار التراكمي للمئين المئوي حسابه بين مجموعة التكرارات التراكمية ، ويمكن توضيح كيفية حساب المئينات من خلال الأمثلة التالية : -

أولاً: حساب المئينات في حالة البيانات الخام غير المجدولة:

لحساب المئينات في حالة البيانات الاحصائية غير المجدولة في فئات يمكن اتباع الخطوات التالية :

1- ترتيب البيانات تصاعدياً ابتداءً من القيمة الدنيا في التوزيع وانتهاءاً بالقيمة العليا .

2- رصد التكرارات المقابلة للبيانات الاحصائية .

3- تحديد التكرارات التراكمية المناظرة للبيانات الاحصائية .

4- تحديد الحدود الفعلية للبيانات الاحصائية .

5- حساب التكرار التراكمي المقابل للمئين المراد حسابه .

6- تحديد موقع التكرار التراكمي المقابل للمئين المراد حسابه في التوزيع .

ثم يصار إلى حساب القيمة التي تناظره في عمود الحدود الفعلية بطريقة النسبة والتناسب .

ومن أجل توضيح الإجراءات السابقة يمكن الاستعانة بالمثال التالي :

مثال (3:1) : إذا كان لديك علامات ثلاثين طالباً على امتحان العلوم موزعة على النحو المبين أدناه : المطلوب حساب المئين (60) لهذه البيانات .

16	16	16	16	15	14	12	13	13	12
20	20	18	18	18	18	18	18	17	17
23	22	21	21	21	21	21	21	21	20

الحل:

الدرجات	التكرار	التكرار التراكمي	الحدود الفعلية
12	2	2	12.5
13	2	4	13.5
14	1	5	14.5
15	1	6	15.5
16	4	10	16.5
17	2	12	17.5
18	6	18	18.5
20	3	21	20.5
21	7	28	21.5
22	1	29	22.5
23	1	30	23.5
المجموع	30		

(المئين 25)

(المئين 60)

(المئين 90)

1- نرتب هذه البيانات تصاعدياً كما هو مبين في الجدول التالي

2- نرصد التكرارات المناظرة لكل درجة من الدرجات وذلك كما يظهر في العمود الثاني من الجدول .

3- نحدد التكرار التراكمي لكل درجة ، وهو بمثابة التكرار المقابل لكل درجة مضافاً إليه التكرار التراكمي المناظر للدرجة التي تسبقها وذلك كما يظهر في العمود الثالث .

4- نجد الحدود الفعلية العليا لكل درجة كما يظهر في العمود الرابع .

5- نحسب التكرار التراكمي المقابل للمئين المراد حسابه من خلال استخدام المعادلة التالية :

$$\text{التكرار التراكمي المقابل للمئين} = \frac{\text{قيمة المئين}}{100} \times \text{مجموع التكرار}$$

أي : $\frac{60}{100} \times 30 = 18$

6- ننظر إلى عمود التكرارات التراكمية لتحديد موقع التكرار التراكمي المقابل للمئين المراد حسابه.

ونلاحظ أنه في المثال الذي نحن بصدده، إن قيمة التكرار التراكمي المقابل للمئين (60) هو 18 وهذه القيمة موجودة في عمود التكرارات التراكمية، وبذلك تكون الدرجة المقابلة لهذا المئين هي الحد الفعلي الأعلى المقابل للتكرار التراكمي (18)، وتساوي 18.5 وذلك كما هو مشار إليه بالسهم في الجدول السابق.

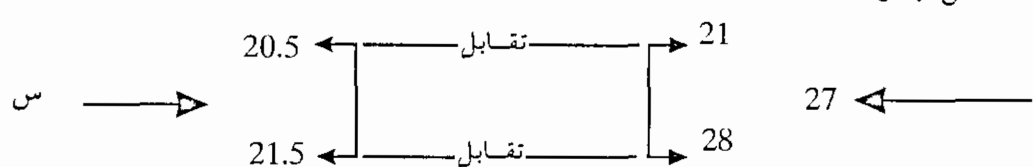
مثال (3:2): في ضوء البيانات الواردة في المثال السابق، احسب المئين 90 والمئين 25.

أ- حساب المئين 90

- التكرار التراكمي المقابل للمئين 90 = $\frac{90}{100} \times 30 = 27$

- بالنظر إلى عمود التكرارات التراكمية في الجدول، نلاحظ أن هذا التكرار يقع بين التكرار التراكمي (21) والمناظر للحد الفعلي (20.5) والتكرار التراكمي (28) المناظر للحد الفعلي (21.5) وذلك كما هو مشار إليه بالسهم في الجدول.

- من الجدول نلاحظ أن



- لتحديد قيمة (س) وهي القيمة المناظرة للتكرار التراكمي 27 نستخدم النسبة والتناسب على النحو الآتي:

(28 - 21) = 7 وهي	تقابل	(21.5-20.5) =1
(27 - 21) = 6 وهي	تقابل	س

وبالتالي نحصل على ما يلي:

7	تقابل	1
6	تقابل	س

- نستخدم النسبة والتناسب لتحديد قيمة (س) من خلال الضرب التبادلي، وذلك على النحو الآتي:

$6 \times 1 = 7 \times$ س

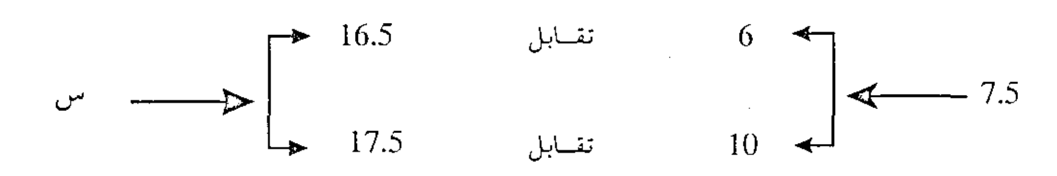
∴ س = $6 \div 7 = 0.88$

- نضيف هذا الناتج إلى الحد الفعلي الأعلى المقابل للتكرار التراكمي الذي يقع دون التكرار التراكمي للمئين 90، أي 0.88 + 20.5 ويساوي (21.38) وهي الدرجة التي تقابل المئين (90). وهكذا يمكن القول أن 90% من الدرجات في التوزيع تقع دون هذه الدرجة.

ب- حساب المئين 25

نتبع نفس الخطوات السابقة في حساب المئين 90 وذلك على النحو الآتي:

$$\frac{25}{100} \times 30 = 7.5$$



4 ←تقابل— 1

بالضرب التبادلي

1.5 ←تقابل— س

∴ 4 س = 1.5

س = $\frac{1.5}{4} \approx 0.38$ تقريبا

∴ المئين 25 = $0.38 + 16.5 = 16.88$

مثال (3 : 3) : أمامك البيانات المبينة في الجدول أدناه، احسب المئين 40، المئين 75، المئين 60:

الدرجات	التكرار	التكرار التراكمي	الحدود الفعلية	
20	2	2	20.5	
23	3	5	23.5	
29	2	7	29.5	
30	4	11	30.5	
32	5	16	32.5	← المئين 40
37	6	22	37.5	← المئين 60
39	3	25	39.5	
42	5	30	42.5	← المئين 75
45	7	37	45.5	
48	2	39	48.5	
50	1	40	50.5	
المجموع	40			

أ- المئين 40: التكرار التراكمي المقابل للمئين 40 هو:

$$16 = 40 \times \frac{40}{100}$$

بالنظر إلى عمود التكرارات التراكمية نلاحظ أن التكرار التراكمي (16) موجود في ذلك العمود، وعليه تكون الدرجة المقابلة للمئين 40 هي 32.5 وذلك كما هو مبين في الجدول أعلاه.

ب- المئين 75

التكرار التراكمي المقابل للمئين 75 هو:

$$30 = 40 \times \frac{75}{100}$$

بالنظر إلى الجدول، نلاحظ أن هذا التكرار أيضاً موجود في عمود التكرارات التراكمية وهو يقابل الحد الفعلي 42.5، وبذلك فإن الدرجة التي تقابل المئين 75 هي 42.5.

ج- المئين 60

التكرار التراكمي المقابل للمئين 60 هو:

$$\frac{60}{100} \times 40 = 24$$

بالنظر إلى الجدول، نلاحظ أن هذا التكرار يقع بين التكرار التراكمي 22 والتكرار التراكمي 25 والحدود الفعلية المناظرة لها على النحو الآتي:

24 →	22	تقابل	37.5	←	س
	25	تقابل	39.5		

وعليه فإن	3	تقابل	2
	2	تقابل	س

وهكذا فإن 3 س = 2

$$\therefore س = \frac{2}{3}$$

س = 0.67

وعليه فالمئين 60 = 0.67 + 37.5 = **38.17**

ثانياً: حساب المئينات في حالة البيانات المجدولة.

لحساب المئينات في حالة البيانات المجدولة نستخدم نفس الإجراءات المتبعة في حسابه في حالة البيانات غير المجدولة ولكن الفرق الوحيد بين هاتين الحالتين هو أن البيانات يتم تبويبها في فئات حسب طول فئة معين في حالة البيانات المجدولة. ولتوضيح ذلك نأخذ المثال التالي:

مثال (3:4): أمامك درجات خمسين طالباً على اختبار ذكاء والمطلوب تلخيصها في فئات وحساب المئينات 25. 30. 50. 80.

الدرجات هي :

63, 56, 52, 52, 51, 45, 45, 78, 76, 74, 73, 70, 70, 70,
69, 65, 65, 90, 90, 89, 87, 85, 85 , 85, 82, 82, 80,
104, 104, 98, 96, 96, 94, 94, 90, 90, 118, 115, 110,
110, 110, 110, 109, 105, 125, 120, 120, 120, 120, 119.

الحل:

أعلى قيمة في التوزيع هي 125

أدنى قيمة في التوزيع هي 45

المدى = 125 - 54 = 80 (المطلوب تشكيل 8 فئات)

طول الفئة = $\frac{80}{8} = 10$

واعتماداً على طول الفئة هذه يتم تحديد الفئات لهذه البيانات وذلك كما هو مبين في الجدول التالي :

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية
49 - 40	2	2	49.5
59 - 50	4	6	59.5
69 - 60	4	10	69.5
79 - 70	7	17	79.5
89 - 80	8	25	89.5
99 - 90	9	34	99.5
109 -100	4	38	109.5
119 - 110	7	45	119.5
129 - 120	5	50	129.5
المجموع	50		

12.5 (المئين 25)
15 (المئين 30)
(50) المئين
40 (المئين 80)

أ- حساب المئين 25

التكرار التراكمي المقابل للمئين 25 = $\frac{25}{100} \times 50 = 12.5$

بالنظر إلى عمود التكرارات التراكمية في الجدول، نلاحظ أن هذا التكرار يقع بين التكرارين التراكميين (10) و (17) والمناظرين للحدود الفعلية (69.5) و (79.5) وذلك كما هو مشار إليه بالسهم أعلاه.

وعليه نجد أن

12.5 →	10	تقابل	69.5	← س
	17	تقابل	79.5	

وهكذا فإن

7	تقابل	10
2.5	تقابل	س

وبالضرب التبادلي نجد أن 7 س = 25

∴ س = $\frac{25}{7}$ = 3.57 تقريباً

إذاً المئين 25 = 3.57 + 69.5 = 73 تقريباً

ب- المئين 30

التكرار التراكمي المقابل للمئين 30 = $\frac{30}{100} \times 50 = 15$

بالنظر إلى عمود التكرارات التراكمية في الجدول نلاحظ أن هذا التكرار يقع بين التكراريين التراكمين (10) و (17) واللذين يقابلان الحدود الفعلية (69.5) و (79.5) وذلك كما هو مشار إليه بالسهم.

وعليه نجد أن

15 →	10	تقابل	69.5	← س
	17	تقابل	79.5	

وهكذا فإن

7	تقابل	10
5	تقابل	س

وبالضرب التبادلي نجد أن:

7 س = 50

س = $\frac{50}{7}$ = 7.14 تقريباً

وعليه فإن المئين 30 = 7.14 + 69.5 = 76.64 تقريباً

ج- المئين 50

التكرار التراكمي للمئين 50 = $\frac{50}{100}$ × 50 = 25

بالنظر إلى عمود التكرارات التراكمية في الجدول، نجد هذا التكرار موجود، وعليه فإن الدرجة المناظرة لهذا المئين هي الحد الفعلي المقابل للتكرار التراكمي (25) ، وبذلك تكون الدرجة 89.5 .

د- المئين 80

التكرار التراكمي للمئين 80 = $\frac{80}{100}$ × 50 = 40

بالنظر إلى عمود التكرارات التراكمية في الجدول، نجد أن هذا التكرار يقع بين التكرارين التراكميين (38) و (45) والمناظران للحدود الفعلية (109.5) و (119.5) وذلك كما هو مشار إليه بالسهم في الجدول.

وعليه نجد أن 38 تقابل 109.5

40 ← س

45 تقابل 119.5

وهكذا نجد أن 7 تقابل 10

2 تقابل س

وبالضرب التبادلي نجد أن 7 س = 20

س = $\frac{20}{7}$ = 2.86 تقريباً

وعليه فإن المئين 80 = 2.86 + 109.5 = 112.36 تقريباً

مثال (3:5) : في ضوء البيانات الواردة في الجدول أدناه احسب المئينات 45 ,65 ,85 .

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية
36 - 30	1	1	36.5
43 - 37	4	5	43.5
50 - 44	3	8	50.5
57 - 51	6	14	57.5
64 - 58	7	21	64.5
71 - 65	3	24	71.5
78 - 72	5	29	78.5
85 - 79	1	30	85.5
المجموع	30		

13.5 (المئين 45) ←

19.5 (المئين 65) ←

25.5 (المئين 85) ←

أ- المئين 45

التكرار التراكمي للمئين 45 = $\frac{45}{100} \times 30 = 13.5$

بالنظر إلى عمود التكرارات التراكمية في الجدول نلاحظ أن هذا التكرار يقع بين التكرارين التراكميين (14)و(8) والمقابلين للحدود الفعلية 50.5، 57.5 وذلك كما هو مشار إليه بالسهم في الجدول.

وعليه نجد أن 8 تقابل 50.5

13.5 ← س

14 تقابل 57.5

وهكذا نجد أن 6 تقابل 7

5.5 تقابل س

وباستخدام الضرب التبادلي نجد أن 6 س = 38.5

وأن س = $\frac{38.5}{6}$ = 6.42 تقريباً

وعليه فإن المئين 45 = 6.42 + 50.5 ≈ 56.92 تقريباً

ب- المئين 65

التكرار التراكمي المقابل للمئين 65 = $\frac{65}{100} \times 30 = 19.5$

بالنظر إلى عـمـود التكرارات التـراكـمـيـة في الجـدول نلاحظ أن هذا التكرار يقع بين التكرارين التراكميين (14) و (21) واللذين يقابلان الحدود الفعلية (57.5) و(64.5) .

وعليه نجد أن 19.5 ⟵ 14 تقابل 57.5 ⟶ س
21 تقابل 64.5

وهكذا نجد أن 7 تقابل 7

5.5 تقابل س

وباستخدام الضرب التبادلي نجد أن

7 س = 38.5

وأن س = $\frac{38.5}{7} = 5.5$

وعليه فإن المئين 65 = 5.5 + 57.5 = 63

ج- المئين 85

التكرار التراكمي المقابل للمئين85 = $\frac{85}{100} \times 30 = 25.5$

وبالنظر إلى عمود التكرارات التراكمية في الجدول نجد أن هذا التكرار يقع بين التكرارين التراكميين (24) و (29) المقابلين للحدود الفعلية (71.5) و (78.5) وذلك كما هو مشار إليه بالسهم في الجدول.

وعليه نجد أن 25.5 ⟵ 24 تقابل 71.5 ⟶ س
29 تقابل 78.5

وهكذا نجد أن 5 تقابل 7

1.5 تقابل س

وبالضرب التبادلي نجد أن 5 س = 10.5

وأن س $= \frac{10.5}{5} = 2.1$

وعليه فإن المثين 85 $= 2.1 + 71.5 = 73.6$

الرتبة المئينية Percentile Rank:

تعرَّف الرتبة المئينية على أنها نسبة مئوية تقابل درجة معينة في توزيع معين. ومثل هذه الرتبة قد تأخذ أي قيمة تتراوح بين صفر و 100% . فمثلاً قد نجد أن درجة طالب في امتحان ما مقدارها 75 والرتبة المئينية 68% بحيث يعني أن 68% من الدرجات تقع دون الدرجة 75. يرمز للرتبة المئينية بالرمز (ي ر) ويتطلب حسابها الخطوات التالية:

1- تحديد التكرارات التراكمية والحدود الفعلية للبيانات.

2- النظر إلى عمود الحدود الفعلية في جدول البيانات وتحديد موقع العلامة (الدرجة) المراد ايجاد الرتبة المئينية لها.

3- استخدام النسبة والتناسب لحساب التكرار التراكمي المقابل للدرجة.

4- قسمة التكرار التراكمي المحسوب على مجموع التكرارات وضرب الناتج بـ 100 % ليمثل الناتج الرتبة المئينية المطلوبة. ولتوضيح ذلك يمكن النظر إلى المثال التالي: -

مثال (3:6): احسب الرتبة المئينية للدرجة (5) في التوزيع التالي:

الدرجات	التكرار	التكرار التراكمي	الحدود الفعلية
1	2	2	1.5
2	3	5	2.5
3	5	10	3.5
4	4	14	4.5
5	6	20	5.5
6	2	22	6.5
7	1	23	7.5
8	2	25	8.5
المجموع	25		

5 ⟶ (بين 4.5 و 5.5)

الحل : نحدد موقع الدرجة (5) في عمود الحدود الفعلية ونلاحظ أنها تقع بين الحدين الفعليين (4.5) و (5.5) واللذين يقابلان التكرارين التراكميين (14) و (20) وذلك كما هو مشار إليه بالسهم في الجدول أعلاه.

وعليه نجد أن 14 تقابل 4.5

س ← 5

20 تقابل 5.5

وهكذا نلاحظ أن 6 تقابل 1

س تقابل $\frac{1}{2}$

وبالضرب التبادلي نجد أن س = $\frac{1}{2} \times 6$

س = 3

وبذلك يكون التكرار التراكمي المطلوب هو $14 + 3 = 17$

وعليه فإن الرتبة المئينية للعلامة (5) = $\frac{17}{25} \times 100 = 68$ %

وهذا يعني بالطبع أن 68% من الدرجات هي دون الدرجة (5).

مثال (3:7): في ضوء البيانات الواردة في الجدول أدناه احسب الرتب المئينية للدرجات (62, 58, 40, 30)

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية	
24 - 20	3	3	24.5	
29 - 25	1	4	29.5	← 30
34 - 30	5	9	34.5	
39 - 35	2	11	39.5	← 40
44 - 40	6	17	44.5	
49 - 45	7	24	49.5	
54 - 50	6	30	54.5	← 58
59 - 55	2	32	59.5	← 62
64 - 60	8	40	64.5	
المجموع	40			

أ- الرتبة المئينية للدرجة 30.

بالنظر إلى عمود الحدود الفعلية في الجدول أعلاه نلاحظ أن هذه الدرجة تقع بين الحدين الفعليين (29.5) و (34.5) واللذين يقابلان التكرارين التراكميين (4) و (9) كما هو مشار إليه أعلاه.

وعليه نجد أن 4 تقابل 29.5

س ← → 30

9 تقابل 34.5

وبالتالي نجد أن 5 تقابل 5

س تقابل $\frac{1}{2}$

وباستخدام الضرب التبادلي نجد أن 5 س = 2.5

وأن س = $\frac{1}{2}$

وبذلك يكون التكرار التراكمي المطلوب هو $4 + \frac{1}{2} = 4\frac{1}{2}$

وبذلك تكون الرتبة المئينية على النحو الآتي

$$\frac{4.5}{40} \times 100 \approx 11\ \%$$

ب- الرتبة المئينية للدرجة 40

نلاحظ أن هذه الدرجة تقع بين الحدين الفعليين (39.5) و (44.5) والمناظرين للتكرارين التراكميين (11) و (17) (انظر إلى الجدول أعلاه).

وبذلك نلاحظ أن 11 تقابل 39.5

س ← → 40

17 تقابل 44.5

أي أن 6 تقابل 5

س تقابل $\frac{1}{2}$

وبالضرب التبادلي نجد أن 5س = 3 : س= $\frac{3}{5}$ = 0.6

وأن التكرار التراكمي المطلوب = 11 + 0.6 =11.6

وبذلك تكون الرتبة المئينية للدرجة 40 هي : $\frac{11.6}{40}$ × 100% =29%

-وهذا يشير إلى أن 29% من الدرجات في ذلك التوزيع تقع دون الدرجة (40).

ج- الرتبة المئينية للدرجة 58.

في الجدول أعلاه، نجد أن هذه الدرجة تقع بين الحدين الفعليين (54.5) و (59.5) المقابلين للتكرارين التراكميين (30) و (32) (انظر إلى الجدول).

وبذلك نجد أن 30 تقابل 54.5

س ⟵ ⟶ 58

32 تقابل 59.5

وعليه نجد أن 2 تقابل 5

س تقابل 3.5

وبالضرب التبادلي نجد أن: 5 س =7

وأن س = $\frac{7}{5}$ = 1.4

وهكذا فإن التكرار التراكمي المطلوب =1.4 + 30 = 31.4

وبذلك فإن الرتبة المئينية للدرجة (58) هي :

$$\frac{31.4}{40} \times 100\ \% = 78.5\ \%$$

د- الرتبة المئينية للدرجة 62.

بالنظر إلى عمود الحدود الفعلية في الجدول أعلاه نجد أن هذه الدرجة تقع بين الحدين الفعليين (59.5)و (64.5) و المقابلين للتكرارين التراكميين (32) و (40).

أي أن س ← 32 تقابل 59.5 → 62

40 تقابل 64.5

وبذلك نجد أن 8 تقابل 5

س تقابل 2.5

وبالضرب التبادلي نجد أن 5 س = 20

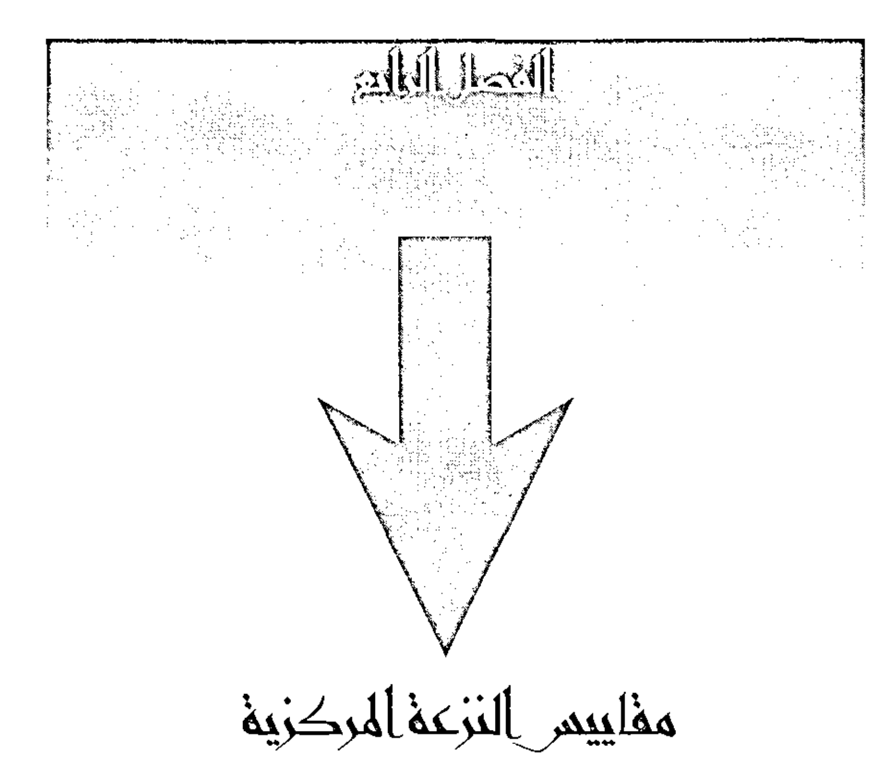
وأن س $= \frac{20}{5} = 4$

وعليه ، فإن التكرار التراكمي المطلوب = 32 + 4 = 36

وأن الرتبة المئينية للدرجة 62 =

$$\frac{36}{40} \times 100\ \% = 90\ \%$$

وهكذا نلاحظ أن 90% من الدرجات في التوزيع أعلاه تقل عن الدرجة 62.



الفصل الرابع

مقاييس النزعة المركزية

Measures of Central Tendency

تمهيد:

بعد الانتهاء من جمع البيانات الاحصائية حول متغير ما وتبويبها وتلخصيها في جداول وتمثيلها بيانياً، قد يتطلب الأمر إيجاد مؤشر احصائي يستخدم لوصف تلك البيانات على نحو أدق. وذلك لأن التمثيل البياني بالرغم أنه يعطي وصفاً سريعاً ومبسطاً للبيانات، إلا أنه يعد غير كافٍ لوصف طبيعة توزيع تلك البيانات.

ففي الكثير من الأحيان يتطلب الأمر إجراء بعض المعالجات الاحصائية على البيانات الكمية مثل تلخيصها في قيمة واحدة تعبر عنها وذلك بهدف تحديد مستوى متغير معين ضمن مجموعة ما، أو مقارنة مستوى هذا المتغير في تلك المجموعة وفق تسلسل زمني، أو مقارنة هذا المتغير مع متغيرات أخرى لنفس المجموعة، أو مقارنة مستواه في عدد من المجموعات. ففي مثل هذه الحالات يستدعي الأمر حساب قيمة معينة ليتم في ضوءها إجراء مثل تلك المقارنات.

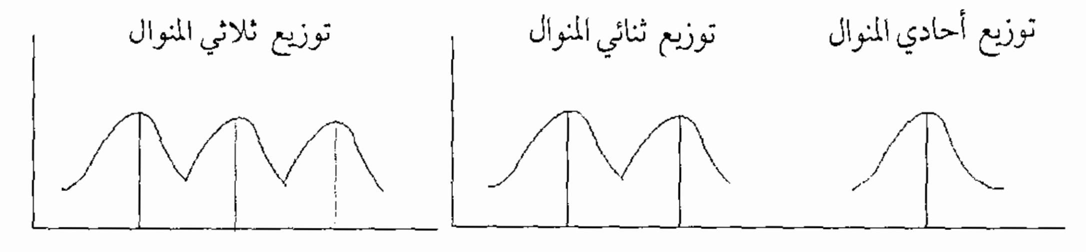
فبالإضافة إلى التمثيل البياني للبيانات، يمكن اللجوء إلى نوعية من المعالجات الاحصائية لوصف تلك البيانات وهي مقاييس النزعة المركزية ومقاييس التشتت، وكل نوع من هذه المعالجات يعطينا وصفاً معيناً عن كيفية توزيع البيانات. وفي هذا الفصل سوف يتم التعرض إلى مقاييس النزعة المركزية، في حين سيتم التعرض في الفصل اللاحق إلى مقاييس التشتت.

تعرف مقاييس النزعة المركزية باسم المتوسطات المركزية، ومثل هذه المقاييس تعطي فكرة عن مستوى المتغير في التوزيع من حيث تحديد مراكز تجمع البيانات لذلك المتغير، وتشمل كل من الوسط والوسيط والمنوال، وفيما يلي عرض لمثل هذه المقاييس.

المنوال The Mode

يُعَرَّف المنوال بأنه المشاهدة أو القيمة الأكثر شيوعاً (تكراراً) في توزيع معين. وفي حالة البيانات المبوبة في توزيعات تكرارية يعد المنوال مركز الفئة الأكثر تكراراً. يعد المنوال من أبسط مقاييس النزعة المركزية وذلك لأن إيجاده لا يتطلب إجراء عمليات إحصائية معقدة، وإنما يتطلب الأمر النظر إلى توزيع البيانات ثم تحديد القيمة أو مركز الفئة الأكثر تكراراً لتكون المنوال لذلك التوزيع. وكمؤشر إحصائي يعد المنوال ذو قيمة كبرى في حالة المتغيرات الاسمية (النوعية أو التصنيفية)، إذ من خلاله يمكن تصنيف وترتيب هذه المتغيرات. أما في حالة المتغيرات الكمية فيكاد يكون أقل استخداماً لأنه لا يدخل في أية تحليلات إحصائية أخرى تقع خارج نطاق وصف البيانات، بالإضافة إلى أنه عرضة للتأثر بحجم البيانات وتغيُّر طول الفئة، ولكنه بالوقت نفسه لا يتأثر بالقيم المتطرفة التي توجد في توزيع معين.

يمكن أن يكون للتوزيع الواحد أكثر من منوال واحد. ففي حالة وجود منوال واحد للتوزيع عندها يسمى بالتوزيع أحادي المنوال، أما إذا كان منوالين لذلك التوزيع فيسمى بالتوزيع ثنائي المنوال، في حين إذا كان ثلاثة منوالات، فعندها يسمى بالتوزيع ثلاثي المنوال وذلك كما هو موضح بالشكل (4:1).



شكل (4:1): أشكال مختلفة من التوزيعات التكرارية

ايجاد المنوال:

يتطلب تحديد المنوال لتوزيع ما، النظر إلى المشاهدات أو الفئات وتعيين المشاهدة أو الفئة الأكثر تكراراً في ذلك التوزيع وذلك حسب طبيعة البيانات التي يتم التعامل معها. وفيما يلي عرض لبعض حالات إيجاد المنوال تبعاً لنوعية البيانات الاحصائية.

أولاً: في حالة المتغيرات الاسمية

لتحديد المنوال في حالة تلك المتغيرات يستلزم الأمر تعيين المتغير الاسمي الذي يتكرر أكثر من غيره من المتغيرات الأخرى. وفي حالة وجود عدة مستويات للمتغير الواحد، يستدعي الأمر تعيين أي المستويات الأكثر تكراراً. ولتوضيح ذلك يمكن النظر إلى الأمثلة التالية:

مثال (4:1) : قام طبيب بفحص زمرة الدم لطلاب فصل دراسي معين عدد أفراده (50) طالباً وكانت نتائج الفحص على النحو الآتي :

زمرة الدم	عدد الأفراد
A	10
B	12
AB	15
O-	05
O+	08

بالنظر إلى البيانات الواردة في المثال السابق ، نجد أن المنوال لهذا التوزيع هو زمرة الدم (AB) وذلك لأن الطلبة الذين يحملون زمرة الدم هذه هم الأكثر عدداً (تكراراً) ، حيث يبلغ عددهم (15) طالباً .

مثال (4:2) : في إحصائية لعدد الحضور لحفلة غنائية كانت الأعداد من كلا الجنسين على النحو الآتي 210 إناث ، و 180 ذكوراً . في هذا المثال نجد أن الإناث هي المنوال لهذا التوزيع لأن عددهن هو الأكثر في الحفلة الغنائية .

مثال (4:3) : في إحصائية لعدد العاملين في شركة ما ، كان توزيع العاملين فيها حسب الوظائف التي يشغلونها على النحو الآتي :

	عدد الأفراد
مهندسون	4
فنيون	120
صيانة	60
تسويق	12
خدمات	33
إداريون	20

نلاحظ أن المنوال لهذا التوزيع هم الفنيون وذلك لأن عددهم (120) موظفاً لانهم الأكثر تكراراً (عداداً) بين مجموع الموظفين .

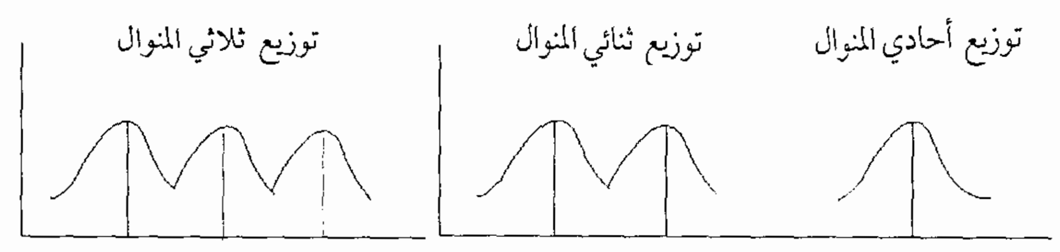
ثانياً : في حالة المتغيرات الكمية غير المبوبة

يستلزم تحديد المنوال في حالة المتغيرات الكمية غير المبوبة النظر فقط إلى توزيع البيانات وتعيين المشاهدة أو القيمة الأكثر تواتراً في ذلك التوزيع لاعتبارها على أنها المنوال ، وذلك كما هو موضح في الأمثلة التالية :

المنوال The Mode

يُعَّرف المنوال بأنه المشاهدة أو القيمة الأكثر شيوعاً (تكراراً) في توزيع معين. وفي حالة البيانات المبوبة في توزيعات تكرارية يعد المنوال مركز الفئة الأكثر تكراراً. يعد المنوال من أبسط مقاييس النزعة المركزية وذلك لأن إيجاده لا يتطلب إجراء عمليات إحصائية معقدة، وإنما يتطلب الأمر النظر إلى توزيع البيانات ثم تحديد القيمة أو مركز الفئة الأكثر تكراراً لتكون المنوال لذلك التوزيع. وكمؤشر إحصائي يعد المنوال ذو قيمة كبرى في حالة المتغيرات الاسمية (النوعية أو التصنيفية)، إذ من خلاله يمكن تصنيف وترتيب هذه المتغيرات. أما في حالة المتغيرات الكمية فيكاد يكون أقل استخداماً لأنه لا يدخل في أية تحليلات إحصائية أخرى تقع خارج نطاق وصف البيانات، بالإضافة إلى أنه عرضة للتأثر بحجم البيانات وتغيّر طول الفئة، ولكنه بالوقت نفسه لا يتأثر بالقيم المتطرفة التي توجد في توزيع معين.

يمكن أن يكون للتوزيع الواحد أكثر من منوال واحد. ففي حالة وجود منوال واحد للتوزيع عندها يسمى بالتوزيع أحادي المنوال، أما إذا كان منوالين لذلك التوزيع فيسمى بالتوزيع ثنائي المنوال، في حين إذا كان ثلاثة منوالات، فعندها يسمى بالتوزيع ثلاثي المنوال وذلك كما هو موضح بالشكل (4:1).



شكل (4:1): أشكال مختلفة من التوزيعات التكرارية

ايجاد المنوال:

يتطلب تحديد المنوال لتوزيع ما، النظر إلى المشاهدات أو الفئات وتعيين المشاهدة أو الفئة الأكثر تكراراً في ذلك التوزيع وذلك حسب طبيعة البيانات التي يتم التعامل معها. وفيما يلي عرض لبعض حالات إيجاد المنوال تبعاً لنوعية البيانات الاحصائية.

أولاً: في حالة المتغيرات الاسمية

لتحديد المنوال في حالة تلك المتغيرات يستلزم الأمر تعيين المتغير الاسمي الذي يتكرر أكثر من غيره من المتغيرات الأخرى. وفي حالة وجود عدة مستويات للمتغير الواحد، يستدعي الأمر تعيين أي المستويات الأكثر تكراراً. ولتوضيح ذلك يمكن النظر إلى الأمثلة التالية:

مثال (4:1) : قام طبيب بفحص زمرة الدم لطلاب فصل دراسي معين عدد أفراده (50) طالباً وكانت نتائج الفحص على النحو الآتي :

زمرة الدم	عدد الأفراد
A	10
B	12
AB	15
O-	05
O+	08

بالنظر إلى البيانات الواردة في المثال السابق ، نجد أن المنوال لهذا التوزيع هو زمرة الدم (AB) وذلك لأن الطلبة الذين يحملون زمرة الدم هذه هم الأكثر عدداً (تكراراً) ، حيث يبلغ عددهم (15) طالباً .

مثال (4:2) : في إحصائية لعدد الحضور لحفلة غنائية كانت الأعداد من كلا الجنسين على النحو الآتي 210 إناث ، و 180 ذكوراً . في هذا المثال نجد أن الإناث هي المنوال لهذا التوزيع لأن عددهن هو الأكثر في الحفلة الغنائية .

مثال (4:3) : في إحصائية لعدد العاملين في شركة ما ، كان توزيع العاملين فيها حسب الوظائف التي يشغلونها على النحو الآتي :

	عدد الأفراد
مهندسون	4
فنيون	120
صيانة	60
تسويق	12
خدمات	33
إداريون	20

نلاحظ أن المنوال لهذا التوزيع هم الفنيون وذلك لأن عددهم (120) موظفاً لانهم الأكثر تكراراً (عداداً) بين مجموع الموظفين .

ثانياً : في حالة المتغيرات الكمية غير المبوبة

يستلزم تحديد المنوال في حالة المتغيرات الكمية غير المبوبة النظر فقط إلى توزيع البيانات وتعيين المشاهدة أو القيمة الأكثر تواتراً في ذلك التوزيع لاعتبارها على أنها المنوال ، وذلك كما هو موضح في الأمثلة التالية :

مثال (4 : 4): في اختبار لفصل دراسي في مادة العلوم كانت علامات الطلاب على النحو الآتي، فما هو المنوال لهذه الدرجات:

20, 11, 12, 15, 17, 17, 17, 17, 16, 16

10, 19, 18, 25, 23, 21, 13, 14, 14, 14

بالنظر إلى الدرجات في التوزيع أعلاه، نلاحظ أن الدرجة (17) هي الأكثر تكراراً، إذ أنها تكررت أربع مرات وبذلك فإن هذه الدرجة هي المنوال لهذا التوزيع.

مثال (4:5): إذا كانت معاملات الذكاء لعدد من الأفراد موزعة على النحو الآتي. فما هو المنوال لهذه المعاملات؟

108, 106, 105, 100, 100, 98, 90, 85

120, 115, 115, 115, 110, 110, 109, 121

121, 121

بالنظر إلى المشاهدات أعلاه، نلاحظ أن لهذا التوزيع منوالين هما (115) و (121) وذلك لأن هاتين الدرجتين تكررت ثلاث مرات.

ثالثاً: في حالة البيانات الكمية المبوبة

يحدد المنوال في حالة البيانات المجدولة من خلال تعيين مركز الفئة الأكثر تكراراً، بحيث تسمى مثل هذه الفئة بالفئة المنوالية، ويسمى مركزها بمنوال التوزيع. وذلك كما هو مبين في الأمثلة التالية:

مثال (4:6): الجدول أدناه يظهر درجات مجموعة من الأفراد عددهم (20) على اختبار للدافعية. عيّن المنوال والفئة المنوالية لذلك التوزيع:

الفئات	التكرار	مركز الفئة
14 - 10	3	12
19 - 15	4	17
24 - 20	7	22
29 - 25	2	27
34 - 30	4	32
المجموع	20	

بالنظر إلى البيانات في الجدول أعلاه نجد أن الفئة المنوالية لهذا التوزيع هي الفئة الثالثة (20 - 24) لأنها الأكثر تكراراً وبذلك يكون مركزها هو المنوال لهذا التوزيع؛ أي أن المنوال لهذا التوزيع = 22 .

مثال (4:7) : يلخص الجدول أدناه عدد ساعات التدريب الأسبوعي التي يقضيها أعضاء فريق رياضي معين في التدريب على مهارة معينة . المطلوب تحديد الفئة المنوالية والمنوال لهذا التوزيع .

الفئات	التكرار	مركز الفئات
11 - 5	4	8
18 - 12	6	15
25 - 19	2	22
32 - 26	6	29
39 - 33	2	36
المجموع	20	

بالنظر إلى البيانات الواردة في الجدول أعلاه، نلاحظ وجود فئتين منواليتين للتوزيع وهي الفئة الثانية (12 - 18) والفئة الرابعة (26 - 32)، وذلك لأن هاتين الفئتين تكررت أكثر من غيرهما، بمعنى أن لهاتين الفئتين أكثر التكرارات وهو (6) . وبذلك يكون لهذا التوزيع منوالان هما مركزا الفئات للفئتين السابقتين وهما على الترتيب (15)و (29) .

ملاحظات عامة حول المنوال

1- المنوال قيمة إحصائية تتمثل في المشاهدة، أو مركز الفئة الأكثر تكراراً في توزيع معين من البيانات الإحصائية .

2- سهولة حسابه بالرغم من وجود بعض الطرق التي تتطلب إجراء بعض العمليات الحسابية لتحديد المنوال مثل طريقة العزوم وطريقة بيرسون، إلا أن الطريقة الأكثر شيوعاً واستخداماً لا تتطلب إجراء أية عمليات حسابية ، وإنما كل ما في الأمر هو النظر إلى التوزيع وتعيين المشاهدة أو مركز الفئة الأكثر تكراراً على اعتبار أنها المنوال .

3- في عملية تحديده لا تدخل جميع القيم (المشاهدات) الواردة في التوزيع، وإنما يقتصر الأمر على مشاهدة واحدة أو بعضاً من المشاهدات . وهذا يعني بالطبع أن المنوال قد يكون مؤشراً مضللاً أو على الأقل غير كافٍ لوصف توزيع معين .

4- كون أن المنوال يعتمد على تكرار المشاهدة أو المشاهدات الأكثر تكراراً، فهو بالتالي لا يتأثر بالقيم المتطرفة في توزيع معين.

فلو نظرنا إلى التوزيع التالي: 1, 15, 16, 16, 16, 16, 17, 18, 87 نلاحظ أن المنوال هو (16) بالرغم من وجود قيمتين متطرفيتين هما (87)، (1).

5- يفضل استخدامه في حالة التوزيعات التكرارية ذات الفئات المفتوحة وذلك لأن حسابه لا يأخذ بعين الاعتبار جميع القيم التي يشتمل عليها التوزيع.

6- في حالة كون غالبية البيانات تتمركز في فئات متباعدة عن بعضها بعضاً فعندها يفضل استخدامه بدلاً من الوسط أو الوسيط.

7- يمكن استخدامه في حالة المتغيرات التي تقاس وفق المقاييس الاسمية أو الرتبية أو الفئوية أو النسبية، ولكن يكون أكثر دلالة في حالة المتغيرات الاسمية.

8- يتأثر المنوال بالتغيرات التي تطرأ على القيم (المشاهدات) في توزيع معين، إذ أنه يتغير تبعاً للتحويلات الخطية التي تجري على القيم. ولتوضيح ذلك يمكن الاستعانة بالمثال التالي:

مثال (4:8): أمامك التوزيع التالي والذي يمثل عدد ساعات التدريب الأسبوعي على مهارة رياضية لعدد من اللاعبين.

10, 9, 8, 7, 6, 6, 6, 5, 5, 4

فما هو المنوال لهذا التوزيع؟

الحل: بالنظر إلى المشاهدات أعلاه، نجد أن العدد (6) هو المنوال لأنه المشاهدة الأكثر تكراراً.

ماذا يحدث لهذا المنوال في حالة إجراء بعض التحويلات الخطية على المشاهدات؟

الحل: في حالة الجمع: إذا تم إضافة مقدار ثابت (س) لجميع المشاهدات في التوزيع، عندها يصبح المنوال الجديد مساوياً للمنوال السابق مضافاً إليه قيمة (س)، أي 6 + س. فمثلاً لو تم إضافة مقدار ثابت قيمته (2) لكل مشاهدة في التوزيع السابق عندها تصبح القيم على النحو الآتي:

12, 11, 10, 9, 8, 8, 8, 7, 7, 6

لاحظ أن المنوال الجديد أصبح = (8) بدلاً من (6) وذلك لأنه تأثر بقيمة الثابت (2) الذي تمت إضافته لجميع القيم.

في حالة الطرح: كما هو الحال في حالة الجمع، فإن المنوال يتأثر بعملية الطرح، فهو يقل بنفس قيمة المقدار الثابت الذي يطرح من كل قيمة بحيث تصبح قيمته الجديدة عبارة عن (المنوال السابق– س).

ففي المثال السابق إذا تم طرح مقدار ثابت فليكن (2) من كل قيمة، عندها تصبح على النحو الآتي:

7, 8, 6, 5, 4, 4, 4, 3, 3, 2

ويصبح المنوال الجديد (4) بدلاً من 6، أي 6 - 2 = 4.

في حالة القسمة: تتأثر قيمة المنوال بناتج قسمة المشاهدات على عدد ثابت ليصبح مساوياً لناتج قسمته على ذلك العدد الثابت، أي قيمة المنوال السابق مقسوما على مقدار الثابت.

وفي المثال السابق يصبح 6 ÷ س. . فإذا تم قسمة جميع القيم الواردة في التوزيع السابق على مقدار ثابت فليكن س = 2، تصبح المشاهدات في ذلك التوزيع على النحو الآتي:

5, 4.5, 4, 3.5, 3, 3, 3, 2.5, 2.5

ويصبح المنوال الجديد (3)، أي 6 ÷ 2 = 3.

في حالة الضرب: يأخذ المنوال الجديد ناتج ضرب المنوال السابق بالمقدار الثابت (س) الذي تضرب فيه جميع القيم في التوزيع، أي يصبح عبارة عن المنوال السابق مضروباً في قيمة المقدار الثابت (س).

ففي حالة مثالنا السابق، تصبح قيمة المنوال الجديد عبارة عن ناتج ضرب المنوال الأول في المقدار الثابت (س)، أي 6 x س.

فلو تم ضرب جميع القيم الواردة في المثال السابق مثال (4:8) بمقدار ثابت فليكن س = 2، عندها تصبح القيم الجديدة في التوزيع على النحو الآتي:

20, 18, 16, 14, 12, 12, 10, 10, 8

وعليه تصبح قيمة المنوال الجديد (12) بدلاً من (6)، أي 6 × 2.

9- في حالة البيانات المجدولة يتأثر المنوال بطول الفئة، حيث أن قيمته تتغير بتغيير طول الفئة. ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:9): لديك البيانات التالية والتي تمثل درجات عشرين طالباً على اختبار ما، وهي موزعة على النحو الآتي:

11, 11, 10, 8, 7, 6, 6, 6, 5, 4, 4

20, 18, 17, 16, 15, 15, 14, 14, 13

المطلوب: تبويب هذه البيانات وفق طول فئة مقداره (3)، ثم وفق طول فئة مقداره (5) وتحديد المنوال في كل حالة، ماذا نلاحظ؟

الحل: في حالة كون طول الفئة مساوياً لـ (3)، فإن توزيع البيانات سوف يكون على النحو الآتي:

الفئات	التكرار	مركز الفئات
5 - 3	3	4
8 - 6	5	7
11 - 9	3	10
14 - 12	3	13
17 - 15	4	16
20 - 18	2	19
المجموع	20	

في ضوء البيانات الواردة في الجدول أعلاه نجد أن الفئة المنوالية الأكثر تكراراً هي الفئة الثانية (6-8) حيث يقابلها خمسة تكرارات وبذلك يكون مركزها الذي تبلغ قيمته (7) المنوال لهذا التوزيع.

أما في حالة كون طول الفئات = 5، فإن التوزيع لهذه البيانات سوف يكون على النحو الآتي:

الفئات	التكرار	مركز الفئات
7 - 3	7	5
12 - 8	4	10
17 - 13	7	15
22 - 18	2	20

نلاحظ أن لهذا التوزيع منوالين وهما مركزاً الفئتين الأكثر تكراراً. حيث نلاحظ أن تكرار الفئة الأول (3 - 7) = 7، كما أن تكرار الفئة الثالثة (13 - 17) = 7. وبذلك يصبح لدينا منوالان هما (5) مركز الفئة الأولى، و (15) مركز الفئة الثالثة.

وهكذا نلاحظ أن قيمة المنوال تتأثر في حالة تغيّر طول الفئة وذلك كما هو موضح في المثال أعلاه.

الوسيط The Median

يعرف الوسيط على أنه القيمة التي تقسم التوزيع إلى قسمين متساويين من حيث عدد المشاهدات، بمعنى آخر هو القيمة التي يقع دونها عدد من الحالات مساوياً لعدد الحالات التي تقع فوقها. وبهذا المنظور فالوسيط هو بمثابة المئين (50)؛ أي الدرجة التي يقع تحتها 50% من المشاهدات وكذلك 50% فوقها. ويعد الوسيط مؤشراً إحصائياً يستخدم أحياناً لوصف تجمع أو تمركز البيانات في توزيع معين، كما أنه يستخدم لإجراء العديد من المعالجات الإحصائية مثل اختبار الفرضيات أو المقارنات الإحصائية بين المجموعات على العديد من المتغيرات ولا سيما في حالة البيانات التي تتطلب معالجات إحصائية لا معلمية (Non Parametric).

حساب الوسيط

يتطلب حساب الوسيط ترتيب البيانات ترتيباً تنازلياً أو تصاعدياً وذلك لأن الوسيط يمثل الدرجة التي تقسم التوزيع إلى قسمين متساويين. وعموماً تختلف طريقة ايجاد الوسيط باختلاف البيانات وذلك على النحو الآتي:

أولاً: حساب الوسيط في حالة البيانات غير المجدولة

لتحديد الوسيط في مثل هذه الحالات يستلزم الأمر ترتيبها تنازلياً أو تصاعدياً ثم تحديد القيمة التي تقع في منتصف توزيع هذه البيانات. ففي حالة كون عدد المشاهدات فردياً، عندها يوجد وسيط واحد لهذا التوزيع يمكن حسابه من خلال المعادلة التالية:

$$\text{الوسيط} = \frac{\text{عددالمشاهدات (ن)} + 1}{2}$$

ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:10): لديك الدرجات التالية

8, 3, 4, 12, 10, 7, 9

احسب الوسيط لهذه القيم:

الحل: 1. يجب ترتيب هذه البيانات إما تصاعدياً أو تنازلياً. فليكن في مثالنا تصاعدياً من القيمة الأدنى إلى الأعلى:

12, 10, 9, ⑧, 7, 4, 3

2- بما أن عدد المشاهدات فردي فالوسيط لها يساوي (عدد المشاهدات + 1 ÷ 2)؛ بذلك فإن القيمة المناظرة للمشاهدة الرابعة في التوزيع هي الوسيط، حيث أنها في مثالنا = (8).

مثال (4:11): احسب الوسيط للتوزيع التالي :

10, 9, 12, 20, 20, 20, 4, 15, 7, 7, 6

الحل : ترتيب البيانات تصاعدياً :

20, 20, 20, 15, 12, 10 , 9, 7, 7, 6, 4

يمكن من خلال العد تحديد القيمة التي تقع في منتصف التوزيع على اعتبار أنها الوسيط، أو من خلال المعادلة التالية :

الوسيط = (ن)+ 1 ÷ 2

عدد المشاهدات (ن)= 11

وعليه فإن رتبة الوسيط هي 11 + 1 ÷ 2 = 6، أي أن القيمة المناظرة للمشاهدة السادسة هي الوسيط لهذا التوزيع، وهي في المثال أعلاه = 10

كما لاحظنا أعلاه أنه في حالة كون عدد المشاهدات في التوزيع فردياً، فسوف يكون هناك وسيط واحد لهذه المشاهدات. ولكن في حالة كون عدد المشاهدات زوجياً نجد أن لهذه المشاهدات وسيطان هما القيم المناظرة لـ $\frac{(ن)}{2}$ و $\frac{(ن)}{2}+1$ ، بحيث يتم اعتبار المتوسط الحسابي لهما على أنه الوسيط، وذلك كما هو موضح بالأمثلة التالية :

مثال (4:12) : أمامك عدد من المشاهدات على النحو الآتي

12, 10, 9, 5, 2, 2

فما هو الوسيط لهذا التوزيع؟

الحل: لاحظ أن عدد المشاهدات في هذا التوزيع عدد زوجي = (6)، وبذلك فإن لهذا التوزيع وسيطين هما القيمة ذات الترتيب الثالث والقيمة ذات الترتيب الرابع، أي :

$\frac{(ن)}{2}$ وهي $\frac{6}{2} = (3)$، $1 + \frac{6}{2} = (4)$

وبالنظر إلى التوزيع أعلاه نلاحظ أن القيمة ذات الترتيب الثالث هي (5) ، أما القيمة ذات الترتيب الرابع فهي (9)، وعليه فإن لهذا التوزيع وسيطين هما على التوالي (5) و (9).

وفي مثل هذه الحالة يتم أخذ المتوسط الحسابي لهاتين القيمتين لتحديد الوسيط وذلك على النحو الآتي:

الوسيط الأول + الوسيط الثاني ÷ 2 ؛ أي 5 + 9 ÷ 2 = (7) وهي قيمة الوسيط لهذا التوزيع.

مثال (4:13) : فيما يلي أطوال 16 طالباً مرتبة تصاعدياً على النحو الآتي:

169, 168, 168, 168, 166, 166,165,

177, 176, 176, 175,173, 172, 172, 172, 170,

احسب المنوال لهذه الأطوال.

الحل : كون أن عدد هذه المشاهدات زوجي فسوف يكون لهذا التوزيع وسيطين وهما المقابلان اللقيمة ذات الترتيب $\frac{(ن)}{2}$ والقيمة ذات الترتيب $\frac{(ن)}{2} + 1$، وذلك على النحو الآتي:

الوسيط الأول: $\frac{16}{2} = 8$: أي المشاهدة الثامنة.

الوسيط الثاني : $\frac{16}{2} + 1 = 9$: أي المشاهدة التاسعة.

وبالنظر إلى التوزيع أعلاه نجد أن قيمة المشاهدة الثامنة هي (170)، في حين قيمة المشاهدة التاسعة هي (172)، وعليه فإن الوسيط لهذا التوزيع يساوي 170 + 172 ÷ 2 = 171.

ثانياً: حساب الوسيط في حالة التوزيعات التكرارية غير المجدولة:

بما أن الوسيط هو المئين 50، فإن عملية حسابه تتطلب تقريباً نفس الإجراءات المستخدمة في حساب المئينات وذلك على النحو الموضح في الأمثلة التالية:

مثال (4:14) : الجدول التالي يظهر عدد الساعات الدراسية لعدد من الطلاب. احسب الوسيط لهذه الساعات.

عدد الساعات	التكرار	التكرار التراكمي	الحدود الفعلية
4	2	2	4.5
6	3	5	6.5
8	5	10	8.5
9	6	16	9.5
10	4	20	10.5
11	3	23	11.5
12	7	30	12.5
المجموع	30		

15 ← (بين 10 و 16)

الحل : أولاً : تحديد القيمة التي تقسم هذا التوزيع إلى قسمين متساويين، أي تحديد القيمة التي تقابل 50 % من التكرار التراكمي، وهي عبارة عن مجموع التكرارات مقسوماً على (2) وذلك على النحو الآتي 30 ÷ 2 = 15. وعليه، فإن هذه القيمة هي التكرار التراكمي الذي يقابل الوسيط . وبالنظر إلى عمود التكرارات التراكمية، نلاحظ أن هذا التكرار يقع بين التكرارين التراكميين (10) و (16) واللذين يقابلان الحدين الفعليين 8.5 و 9.5.

وعليه نجد أن 10 تقابل 8.5

15 ← → س

16 تقابل 9.5

أي أن 6 تقابل 1

5 تقابل س

وبالضرب التبادلي نجد أن 6 س = 5

وأن س 5 ÷ 6 = 0.83

وهكذا فإن قيمة الوسيط = 8.5 + 0.83 = 9.33 تقريباً.

مثال (4:15): يلخص الجدول أدناه درجات عشرين طالباً على أحد الامتحانات . احسب الوسيط لهذه الدرجات .

الدرجة	التكرار	التكرار التراكمي	الحدود الفعلية
12	4	4	12.5
15	5	9	15.5
18	1	⑩	18.5 ←
20	3	13	20.5
22	4	17	22.5
25	2	19	25.5
30	1	20	30.5
المجموع	20		

الحل : القيمة التي تقابل 50 % من التكرار التراكمي هي 20 ÷ 2 = ⑩

بالنظر إلى عمود التكرارات التراكمية في الجدول أعلاه نجد أن هذه القيمة موجودة ، وذلك كما هو مشار إليه بالسهم . وعليه فأن الوسيط لهذا التوزيع هو قيمة الحد الفعلي المقابل لهذا التكرار التراكمي ، أي أن الوسيط = 18.5 .

ثانياً: حساب الوسيط في حالة التوزيعات التكرارية المجدولة.

في الواقع لا تختلف طريقة حساب الوسيط في حالة التوزيعات التكرارية المجدولة عنها في حالة البيانات غير المجدولة ، باستثناء أننا نتعامل مع الحد الفعلي الأعلى للفئات . وعموماً يمكن الاستعانة بالأمثلة التالية لتوضيح كيفية تحديد الوسيط للبيانات المجدولة .

مثال (4:16) : الجدول التالي يلخص أوزان مجموعة من الأفراد ، احسب الوسيط لهذه الأوزان .

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية
44 - 40	1	1	44.5
49 - 45	3	4	49.5
54 - 50	12	16	54.5
59 - 55	8	24	59.5
64 - 60	4	28	64.5
69 - 65	2	30	69.5
المجموع	30		

15 ⟵ (سهم يشير إلى ما بين 4 و 16)

الحل:

1- تحديد القيمة التي تقسم التكرارات التراكمية إلى نصفين متساويين، أي التكرار التراكمي الذي يقابل 50 % من الحالات. وهي في المثال التالي ناتج قسمة مجموع التكرارات على 2؛ أي

30 ÷ 2 = 15.

إذا القيمة (15) هي التكرار التراكمي المقابل للوسيط.

بالنظر إلى عمود التكرارات التراكمية في الجدول، نلاحظ أن هذا التكرار يقع بين التكرارين التراكميين (4) و (16) واللذين يقابلان الحدين الفعليين (49.5) و (54.5)، وذلك كما هو مشار إليه بالسهم.

وعليه نجد أن 4 تقابل 49.5

15 ⟵ ⟶ س

16 تقابل 54.5

وهكذا نجد أن 12 تقابل 5

11 تقابل س

وبالضرب التبادلي (النسبة والتناسب) نجد أن 12 س = 55

س = 55 ÷ 12 = 4.58 تقريباً.

وعليه فإن الوسيط لهذه الأوزان هو 49.5 + 4.58 = 54.08 تقريباً.

مثال (4:17): احسب الوسيط للبيانات الواردة في الجدول أدناه.

الفئات	التكرارات	التكرار التراكمي	الحدود الفعلية
69 - 60	4	4	69.5
79 - 70	2	6	79.5
89 - 80	5	11	89.5
99 - 90	6	17	99.5
109 -100	3	20	109.5
المجموع	20		

10 ⟶ (بين التكرارين التراكميين 6 و 11)

الحل: التكرار التراكمي المقابل للوسيط = 20 ÷ 2= 10. في عمود التكرارات التراكمية نجد أن هذا التكرار يقع بين التكراريين التراكميين والحدين الفعليين المناظران لهما كما هو مبين أدناه

6 تقابل 79.5

10 ⟵ ⟶ س

11 تقابل 89.5

وعليه فإن 5 تقابل 10

4 تقابل س

وباستخدام الضرب التبادلي (النسبة والتناسب) نجد أن

5 س = 40

وأن س = 8

وعليه فإن الوسيط لهذا التوزيع = 79.5 + 8 = 87.5

الخصائص العامة للوسيط

أولاً: يمثل الوسيط قيمة في التوزيع يقع دونها وفوقها 50 % من عدد الحالات. فالوسيط هو نقطة المنتصف للقياسات المختلفة لمتغير ما، وهو ما يعرف بالمئين 50.

ثانياً: بما أن الوسيط هو بمثابة القيمة التي تقسم التوزيع إلى قسمين متساويين، فمثل هذه القيمة لا تتأثر بوجود القيم المتطرفة. ولتوضح ذلك فلنأخذ المثال التالي:

مثال (4:18): أمامك التوزيع التالي لدرجات خمس طلاب على امتحان ما مرتبة ترتيباً تصاعدياً:

21, 20, 17, 15, 14

نلاحظ أعلاه أن قيمة الوسيط لهذه الدرجات هو (17)

ماذا يحدث للوسيط لو كانت درجة الطالب الأول (3)، ودرجة الطالب الخامس (30).

30, 20, 17, 15 , 3

نلاحظ أن قيمة الوسيط بقيت كما هي (17) حيث أنها لم تتأثر بوجود القيم المتطرفة، وعليه فإنه يفضل استخدام الوسيط لوصف البيانات في حالة التوزيعات التي تشتمل على قيم متطرفة.

ثالثاً: : يمكن تحديد الوسيط في حالة عدم توفر بعض البيانات (Missing data) عن بعض الحالات وذلك عندما تتوفر لدينا بعض المعلومات المسبقة عن تلك الحالات. فعلى سبيل المثال إذا سألنا سبعة أشخاص عن أعمارهم وأجاب خمسة منهم على نحو دقيق وامتنع اثنان عن الإجابة، ففي مثل هذه الحالة يمكن تحديد الوسيط في ضوء معرفتنا عن هذين الاثنين اللذين امتنعا عن الإجابة.

فلنفرض أن إجابات الخمسة أشخاص كانت على النحو الآتي:

33, 32, 32, 31, 30

فإذا كان لدينا معرفة أن عمر هذين الاثنين هو أقل من أعمار الأشخاص الآخرين، فعندها يصبح التوزيع على النحو الآتي:

33, 32, 32, (31), 30, 30>, 30>

وبالتالي يكون الوسيط لهذا التوزيع هو (7 + 1 ÷2= 4)، أي العمر المقابل للمشاهدة الرابعة وهو (31).

رابعاً: يستخدم الوسيط في حالة المتغيرات التي تقاس وفق المقاييس الرتبية والفئوية والنسبية، في حين لا يصلح في حالة المتغيرات الاسمية. كما ويستخدم في إجراء العمليات الاحصائية مثل اختبار الفرضيات وإجراء المقارنات في حالة البيانات الاحصائية اللامعلمية.

خامساً: تتأثر قيمة الوسيط بالتحويلات الخطية التي تطرأ على البيانات مثل الجمع والطرح والقسمة والضرب. ففي حالة جمع مقدار ثابت (س) لجميع المشاهدات في توزيع معين، فإن قيمة الوسيط تزداد بمقدار هذا الثابت، كما أن طرح مقدار ثابت (س) من جميع المشاهدات يؤدي إلى انخفاض الوسيط بنفس قيمة المقدار الثابت. أما في حالة الضرب نجد أن قيمة الوسيط تزداد لتساوي حاصل ضرب قيمة الوسيط بالمقدار الثابت، ولكن في حالة القسمة، تصبح قيمة الوسيط هي ناتج قسمته على المقدار الثابت. ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:19): أمامك المشاهدات التالية:

10, 8, 6, 4, 2

الوسيط لهذه المشاهدات هو (6)

ماذا يحدث للوسيط لو تم إضافة مقدار ثابت لكل مشاهدة في التوزيع، فليكن (س = 2)عندها تصبح المشاهدات.

12, 10, 8, 6, 4

وبذلك تصبح القيمة الجديدة للوسيط (8)، حيث أن الوسيط ازداد بنفس قيمة المقدار الثابت الذي تمت إضافته لجميع القيم في التوزيع؛ أي 6 + 2.

ماذا يحدث للوسيط في حالة طرح مقدار ثابت، فليكن (س) = 2 من جميع المشاهدات؟

الجواب: تتناقص قيمة الوسيط بنفس قيمة المقدار الثابت، وذلك على النحو الآتي:

المشاهدات 2 ,4 ,6 ,8 ,10 الوسيط لها = 6

في حالة طرح مقدار ثابت فليكن س = 2 من هذه المشاهدات، عندها تصبح المشاهدات الجديدة على النحو الآتي:

10, 6, 4, 2, 0

وبذلك تصبح قيمة الوسيط الجديد (4)، وهي عبارة عن 6 - 2، أي الوسيط السابق مطروحاً منه قيمة المقدار الثابت.

ماذا يحدث للوسيط في حالة قسمة المشاهدات على مقدار ثابت س = (2)؟ الجواب تصبح قيمة الوسيط عبارة عن ناتج قسمة الوسيط السابق على (2).

المشاهدات 2 ,4 ,6 ,8 ,10 الوسيط لها = 6

في حالة قسمة جميع المشاهدات على مقدار ثابت س = (2)، عندها تصبح المشاهدات الجديدة على النحو الآتي:

5, 4, 3, 2, 1

وبذلك تصبح قيمة الوسيط الجديد (3)، وهي عبارة عن حاصل قسمة الوسيط السابق على المقدار الثابت، أي 6 ÷ 2 = (3).

ماذا يحدث للوسيط في حالة ضرب المشاهدات بمقدار ثابت وليكن س = (2)؟ الجواب: تصبح قيمة الوسيط الجديد مساوية لناتج ضرب الوسيط السابق بالمقدار الثابت.

المشاهدات: 2 ,4 ,6 ,8 ,10 الوسيط لها = 6

في حالة ضرب جميع المشاهدات بمقدار ثابت وليكن س = 2، عندها تصبح المشاهدات الجديدة على النحو الآتي:

20, 16, 12, 8, 4

وبذلك تصبح قيمة الوسيط الجديد (12)، وهذه القيمة هي عبارة عن ناتج (6 x 2)، أي ناتج ضرب الوسيط السابق بالمقدار الثابت.

سادساً: تتأثر قيمة الوسيط في حالة البيانات المجدولة في توزيعات تكرارية بطول الفئة، حيث تتغير قيمته بتغير طول الفئة. فمع زيادة طول الفئة يقل عدد الفئات مما يؤدي بالتالي الى أن تقترب قيمة الوسيط من قيمته في حالة عدم جدولة مثل هذه البيانات. ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:20): فيما يلي درجات 30 طالباً على امتحان المستوى في اللغة الانجليزية.

15, 14, 14, 13, 13, 13, 12, 11, 11, 10, 08, 07, 07

21, 20, 20, 18, 18, 18, 17, 17, 17, 16

27, 25, 24, 24, 23, 22, 21

المطلوب:

أ- حساب الوسيط لهذه الدرجات في حالة عدم جدولتها .

ب- تبويب هذه البيانات في فئات وفق طول فئة (3) وحساب الوسيط لها .

ج- تبويب البيانات في فئات وفق طول فئة مقداره (7)، وحساب قيمة الوسط لها .

د- قارن بين قيم الوسيط في الحالات الثلاث؟

الحل:

أ- قيمة الوسيط في حالة عدم جدولة البيانات:

بما أن عدد القيم زوجي، فإن قيمة الوسيط لهذه البيانات هو عبارة عن المعدل (المتوسط الحسابي) للقيمتين الوسطيتين

وذلك على النحو الآتي: $\frac{(ن)}{2}$ و $\frac{(ن)}{2} + 1$

القيمة الوسيطية الأولى: $\frac{30}{2} = (15)$ المشاهدة الخامسة عشرة

القيمة الوسيطية الثانية: $\frac{30}{2} + 1 = (16)$ المشاهدة السادسة عشرة

وبالنظر إلى التوزيع نجد أن قيمة المشاهدة الخامسة عشرة هي (17). كما أن قيمة المشاهدة السادسة أيضاً هي (17). وعليه فإن الوسيط لهذه الدرجات هو $17 + 17 \div 2 = 17$.

ب- في حالة تبويب هذه البيانات في جداول تكرارية وفق طول فئة مقداره (3).

التكرار التراكمي المقابل لقيمة الوسيط = $30 \div 2 = 15$. وبالنظر إلى عمود التكرارات التراكمية نجد أن هذه القيمة تقع بين التكرارين التراكميين (12) و (17) واللذين يقابلان الحدين الفعليين (14.5) و (17.5) وذلك كما هو موضح في الجدول أدناه .

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية
8 - 6	3	3	8.5
11 - 9	3	6	11.5
14 - 12	6	12	14.5 ← 15
17 - 15	5	17	17.5
20 - 18	5	22	20.5
23 - 21	4	26	23.5
26 - 24	3	29	26.5
29 - 27	1	30	29.5
المجموع	30		

12 تقابل 14.5

15 ← → س

17 تقابل 17.5

وعليه فإن 5 تقابل 3

3 تقابل س

وبالضرب التبادلي (النسبة والتناسب) نجد أن 5 س = 9

$$س = \frac{9}{5} \approx 1.8$$

وعليه فإن قيمة الوسيط = 1.8 + 14.5 = 16.3 .

ج- في حالة تبويب البيانات في جداول تكرارية وفق طول فئة مقداره (7) .

التكرار التراكمي المقابل لقيمة الوسيط = 30 ÷ 2 = 15 ، وبالنظر إلى عمود التكرارات التراكمية في الجدول التالي ، نجد أن هذا التكرار يقع بين التكرارين التراكميين (10) و (22) واللذين يناظران الحدين الفعليين (13.5) و (20.5) وذلك على النحو الآتي :

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية
13 - 7	10	10	13.5
20 - 14	12	22	20.5
27 - 21	8	30	27.5
المجموع	30		

15 ←

10 تقابل 13.5

15 ← → س

22 تقابل 20.5

وعليه فإن 12 تقابل 7

5 تقابل س

وبالضرب التبادلي (النسبة والتناسب) نجد أن: 12 س = 35

س = 35 ÷ 12 = 2.92 تقريباً

وعليه فإن قيمة الوسيط = 13.5 + 2.92 = 16.42 تقريباً

د- قارن بين قيم الوسيط في الحالات الثلاث؟

كانت قيمة الوسيط في الحالة الأولى هي (17)، في حين بلغت قيمته في حالة جدولة البيانات في توزيعات تكرارية وفق طول فئة (3) هي (16.3)، اما في حالة تبويب البيانات في توزيعات تكرارية وفق طول فئة مقداره (7)، أصبحت قيمة الوسيط (16.42). وهكذا نلاحظ أن مع زيادة طول الفئة، تقل عدد الفئات، أي أن البيانات تتمركز في مدى أضيق، مما يؤدي بالتالي إلى أن تقل قيمة الوسيط عن قيمته الأصلية في حالة عدم جدولة البيانات.

المتوسط الحسابي The Mean

يطلق على المتوسط الحسابي اسم الوسط الحسابي "Arithemetic mean" ويعرف بأنه مركز تجمع البيانات في توزيع معين، وذلك لأنه مؤشر لنقطة التوازن لتلك البيانات. إحصائياً يعرّف المتوسط على أنه معدل القيم في توزيع معين ممثلاً ذلك في قسمة مجموع القيم على عددها.

يعد المتوسط الحسابي من أكثر مقاييس النزعة المركزية شيوعاً واستخداماً في وصف البيانات الاحصائية وذلك لأنه أكثرها استقراراً. وهذه الصفة تجعل منه مؤشراً مرغوباً فيه في حالة إجراء العديد من العمليات الاحصائية المتقدمة مثل تحديد فترات الثقة، واختبار الفرضيات وإجراء المقارنات أو أية عمليات إحصائية اخرى تتعلق بالإحصاء التحليلي.

يتأثر المتوسط الحسابي بالقيم المتطرفة في التوزيع وذلك لأن في عمليات حسابه يؤخذ بعين الاعتبار جميع القيم الموجودة في التوزيع، وهذا بالتالي يجعل منه مؤشراً خاطئاً في حالة التوزيعات التكرارية الملتوية، أي التوزيعات التي تتضمن قيماً متطرفة أو غير متجانسة.

حساب المتوسط الحسابي:

عموماً يمثل المتوسط المعدل العام للقيم في التوزيع، أي مجموعها مقسوماً على عددها. وبالرموز

$$\bar{س} = \sum س \div ن$$

$$\text{الوسط} \quad (\bar{س}) = \frac{\sum س}{ن}$$

حيث $\bar{س}$: تشير إلى الوسط

$\sum س$: وهي مجموع القيم الموجودة في التوزيع وتشمل :

$(س_1 ، س_2 ، س_3 \cdots\cdots\cdots س_ن)$

ن : عدد المشاهدات (القيم) في التوزيع.

هذا وتختلف طريقة حساب المتوسط الحسابي باختلاف طبيعة البيانات من حيث كونها بيانات خام أو مجدولة في توزيعات تكرارية. وفيما يلي عرض لطريقة حساب المتوسط الحسابي تبعاً للحالات المختلفة للبيانات.

أولاً: حساب الوسط في حالة البيانات الخام: يتم حساب الوسط في حالة مثل هذه البيانات من خلال المعادلة التالية $(\bar{س}) = \frac{\sum س}{ن}$ وفقاً للخطوات التالية :

1- ايجاد مجموع جميع المشاهدات في التوزيع $\sum س$

2- قسمة المجموع على عدد المشاهدات بحيث يمثل الناتج المتوسط الحسابي لهذه المشاهدات .

مثال (4:21) : احسب المتوسط الحسابي للمشاهدات التالية :

30,　28,　22,　18,　12

الحل : نجد مجموع هذه المشاهدات

$$\sum س = 12 + 18 + 22 + 28 + 30 = 110$$

تقسم مجموع المشاهدات على عددها بحيث يكون الناتج المتوسط الحسابي لها

$$\frac{110}{5} = 22$$ وهكذا فإن

المتوسط الحسابي (سَ) لهذه المشاهدات يساوي (22) .

مثال (4:22) : احسب المتوسط الحسابي للمشاهدات التالية :

-6،　-2،　صفر،　4،　6،　8،　11

الحل :　$(\bar{س}) = \frac{\sum س}{ن}$

$$= \frac{(-6) + (-2) + صفر + 4 + 6 + 8 + 11}{7}$$

$$\frac{21}{7} = 21 = 3$$ وهو المتوسط الحسابي لهذه المشاهدات

ثانياً: حساب المتوسط الحسابي في حالة البيانات التكرارية غير المجدولة :

يتم حساب المتوسط الحسابي في مثل هذه الحالات من خلال استخدام المعادلة التالية :

$$\bar{س} = \frac{\sum س \times ت}{\sum ت} .$$

حيث ت : هو التكرار المقابل لكل مشاهدة في التوزيع ، $\sum ت$: مجموع التكرارات الكلي .

ويمكن اتباع الخطوات التالية في حسابه :

1- ايجاد ناتج ضرب المشاهدة في التكرار المقابل لها (س × ت) .

2- ايجاد المجموع الكلي لناتج ضرب المشاهدات في تكراراتها $\sum$(س × ت) .

3- قسمة المجموع الكلي على مجموع التكرارات بحيث يكون الناتج المتوسط الحسابي لهذه المشاهدات.

مثال (4:23) : الجدول التالي يبيّن درجات مجموعة من الطلاب على امتحان في مادة اللغة العربية. احسب المتوسط الحسابي لهذه الدرجات.

الدرجة (س)	التكرار (ت)	س × ت
14	3	42
15	5	75
16	4	64
17	3	51
18	2	36
19	2	38
20	1	20
المجموع	20	326

الحل : من خلال البيانات الواردة في الجدول أعلاه

نلاحظ أن $\sum$ت = 20

وان $\sum$س × ت = 326

وعليه فإن المتوسط الحسابي لهذه الدرجات = $\frac{326}{20}$ = 16.3 تقريباً

مثال (4:24) : فيما يلي أوزان مجموع من الطلاب، احسب المتوسط الحسابي لهذه الأوزان؟

المشاهدات (س) :

51, 50, 49, 49, 49, 47, 48, 53, 53, 53

52, 51, 51, 51, 56, 55, 55, 54, 54, 54

54, 60, 59, 59, 59, 58, 57, 57, 62, 60

الحل:

أولاً: نلخص هذه البيانات في جدول يظهر التكرارات المقابلة لكل مشاهدة في التوزيع على النحو الآتي:

الوزن (س)	التكرار (ت)	س × ت
47	1	47
48	1	48
49	3	147
50	1	050
51	4	204
52	1	052
53	3	159
54	4	216
55	2	110
56	1	056
57	2	114
58	1	058
59	3	177
60	2	120
62	1	062
المجموع	30	1620

ثانياً: نحدد التكرارات المقابلة لكل قيمة (وزن) وذلك كما هو مبين في العمود الثاني من الجدول.

ثالثاً: نجد مجموع التكرارات $\sum$ت ، وهو يساوي (30) وذلك كما هو مبين في أسفل العمود الثاني.

رابعاً: نجد حاصل ضرب الأوزان بالتكرارات المقابلة لها، ثم نجد المجموع الكلي لها وذلك كما هو مبين في العمود الثالث.

خامساً: نحسب ناتج قسمة المجموع الكلي على مجموع التكرارات ليمثل المتوسط الحسابي، وذلك وفق المعادلة التالية:

$$\bar{س} = \frac{\sum (س \times ت)}{\sum ت} = \frac{1620}{30} = 54$$

ثالثاً: حساب المتوسط الحسابي في حالة البيانات التكرارية المجدولة.

كما هو الحال في البيانات التكرارية غير المجدولة، يمكن استخدام نفس المعادلة لحساب المتوسط الحسابي في البيانات التكرارية لمجدولة ولكن يتم في مثل هذه الحالة استخدام مراكز الفئات بدلاً من القيم الحقيقية في التوزيع. وبذلك تكون المعادلة على النحو التالي:

المتوسط الحسابي = مجموع حاصل ضرب مراكز الفئات × التكرار

وبالرموز : $\bar{س} = \frac{\sum م س \times ت}{\sum ت}$

مثال (4:25): أحسب المتوسط الحسابي للبيانات الواردة في الجدول أدناه.

الفئات	التكرار (ت)	م س	م س × ت
28 - 20	8	24	192
37 - 29	7	33	231
46 - 38	12	42	504
55 - 47	6	51	306
64 - 56	9	60	540
73 - 65	8	69	552
المجموع	50		2325

الحل :

1- نجد مراكز الفئات وهي عبارة عن ناتج قسمة مجموع الحدين الأعلى والأدنى للفئة على اثنين. وذلك كما يظهر في العمود الثالث في الجدول السابق :

2- نجد ناتج ضرب مركز الفئات × التكرار، ثم نجد المجموع الكلي لناتج ضرب مراكز الفئات في التكرارات وذلك كما هو مبين في العمود الرابع في الجدول أعلاه.

3- نقسم المجموع الكلي على مجموع التكرارات وبذلك نحصل على المتوسط الحسابي.

أي المتوسط $\bar{س} = \frac{\sum م_س \times ت}{\sum ت}$

$$\frac{2325}{50} = 46.5$$

هناك طرق أخرى لحساب المتوسطات الحسابية مثل طريقة الترميز (الانحرافات الترتيبية للفئات) وطريقة الوسط الفرضي.

ففي طريقة الوسط الفرضي يتم اختيار أي مشاهدة أو مركز فئة في التوزيع على اعتبار أنه الوسط الفرضي ويضاف إليه مجموع انحرافات المشاهدات عن تلك القيمة مضروباً في تكراراتها. ولتوضيح ذلك فلنأخذ الأمثلة التالية :

مثال (4:26) : باستخدام طريقة الوسط الفرضي، احسب المتوسط الحسابي للبيانات التالية :

80, 45, 30, 20, 15

الحل:

المتوسط الحسابي $(\bar{س}) = و + \frac{\sum (س - و)}{ن}$

حيث و : هو المتوسط الفرضي

$\sum (س - و)$: مجموع انحرافات المشاهدات عن الوسط الفرضي

ن عدد المشاهدات

افترض أنه تم اختيار المشاهدة (30) على اعتبار أنها الوسط الفرضي.

باستخدام المعادلة أعلاه لحساب المتوسط الحسابي الحقيقي نحصل على ما يلي :

$$\bar{س} = 30 + \frac{\big((30-80) + (30-45) + (30-30) + (30-20) + (30-15)\big)}{5}$$

$$= 30 + \frac{\big((50+ 15 + صفر + 10- + 15-)\big)}{5}$$

$$= 30 + \frac{40}{5}$$

= 30 + 8 = 38 وهو الوسط الحقيقي لهذه البيانات

مثال (4:27) : استخدم طريقة الوسط الفرضي لحساب المتوسط الحسابي للبيانات التكرارية غير المجدولة المبينة في الجدول أدناه على اعتبار أن الوسط الفرضي لها = 40

الدرجات	التكرار (ت)	(س - و)	(س - و) ت
15	2	25-	50-
25	3	15-	45-
30	4	10-	40-
45	5	5+	25+
50	6	10+	60+
60	5	20+	100+
المجموع	25		50

الحل : نحسب انحرافات القيم عن الوسط الفرضي وذلك كما هو مبين في العمود الثالث في الجدول أعلاه.

2- نجد مجموع ناتج ضرب انحرافات المشاهدات عن الوسط الفرضي في التكرار، (أي $\sum$(س - و) × ت) وذلك كما هو مبين في العمود الرابع في الجدول السابق، وهو = 50

أي أن $\sum$(س - و) × ت = 50

3- نقسم هذا الناتج على مجموع التكرارات على النحو الآتي:

$$\frac{\sum (س - و) \times ت}{\sum ت} = \frac{50}{25} = 2$$

4- نعوض هذه القيم في المعادلة التالية

$$\bar{س} = و + \frac{\sum (س - و) \times ت}{\sum ت}$$

$$= 40 + \frac{50}{25} = 40 + 2 = 42$$

وبذلك يمثل هذا الناتج المتوسط الحسابي للبيانات .

مثال (4:28) : احسب المتوسط الحسابي للبيانات التكرارية المبوبة في الجدول أدناه على اعتبار أن الوسط الفرضي لها مركز الفئة الرابعة ويساوي 22.

الفئات	التكرار (ت)	م س	(م س - و)	$\sum$(س - و) X ت
9 - 5	2	7	15-	30-
14 - 10	7	12	10-	70-
19 - 15	3	17	5-	15-
24 - 20	4	22	صفر	صفر
29 - 25	5	27	5	25
34 - 30	10	32	10	100
39 - 35	9	37	15	135
المجموع	40			145

الحل:

(1) نحسب انحرافات مراكز الفئات عن الوسط الفرضي (م س - و)، وذلك كما هو مبين في العمود الرابع في الجدول أعلاه .

(2) نحسب مجموع ناتج ضرب انحرافات مراكز الفئات عن الوسط الفرضـي في التـكرار (∑(م س - و) × ت)، وذلك كما هو موضح في العمود الخامس في الجدول.

$$\sum(\text{م س} - \text{و}) \times \text{ت} = 145$$

(3) نقسم هذا الناتج على مجموع التكرارات.

$$\frac{\sum(\text{م س} - \text{و}) \times \text{ت}}{\sum \text{ت}} = \frac{145}{40} \approx 3.6$$

(4) نعوض هذه القيم في المعادلة التالية للحصول على المتوسط الحسابي.

$$\text{المتوسط الحسابي } (\bar{\text{س}}) = \text{و} + \frac{\sum(\text{م س} - \text{و}) \times \text{ت}}{\sum \text{ت}}$$

= 22 + 3.6 = 25.6 وهو المتوسط الحسابي لهذا التوزيع

الوسط الموزون Weighted Mean

يعرف هذا الوسط باسم الوسط المرجح وهو بمثابة وسط الأوساط. ويستخدم عندما تتوفر متوسطات حسابية لمجموعات جزئية لعينة معينة على متغير ما ويراد تحديد المتوسط الحسابي للعينة ككل على ذلك المتغير. ويتم حساب الوسط الموزون من خلال المعادلة التالية:

$$\bar{\text{س}} = \frac{\bar{\text{س}}_1 \text{ن}_1 + \bar{\text{س}}_2 \text{ن}_2 + \ldots \bar{\text{س}}_? \text{ن}_?}{\text{ن}_1 + \text{ن}_2 + \ldots \text{ن}_?}$$

حيث $\bar{\text{س}}_1$: الوسط الحسابي للمجموعة الجزئية الأولى على المتغير

$\bar{\text{س}}_2$: الوسط الحسابي للمجموعة الجزئية الثانية.

$\bar{\text{س}}_3$: الوسط الحسابي للمجموعة الجزئية الأخيرة.

ن_1 ، ن_2 ، ن : اعداد الأفراد في المجموعات الجزئية.

مثال (4:29) : لدى مدرس ثلاثة فصول أجري لها امتحان بمادة الرياضيات وكانت النتائج على النحو الآتي :

الفصل	عدد الطلاب فيه (ن)	المتوسط الحسابي (سَ) لدرجاتهم
أ	30	25
ب	40	28
ج	35	26

احسب الوسط الموزون (الوسط الحسابي) لجميع الفصول على هذا الامتحان؟

الحل:

$$\bar{س} = \frac{\bar{س}_1 ن_1 + \bar{س}_2 ن_2 + \ldots . \bar{س}_3 ن_3}{ن_1 + ن_2 + \ldots . ن_3}$$

$$= \frac{26 \times 35 + 28 \times 40 + 25 \times 30}{35 + 40 + 30}$$

$$= \frac{910 + 1120 + 750}{105} = \frac{2780}{105}$$

$= 26.48$ تقريباً

حساب الوسط الحسابي للتوزيعات التكرارية ذات الفئات المفتوحة:

في بعض الأحيان نواجه توزيعات تكرارية تشتمل على فئات مفتوحة سواءً في بدايتها أو في مؤخرتها ، وفي مثل هذه الحالة يصعب اتباع الخطوات السابقة لايجاد المتوسط الحسابي وذلك لعدم امكانية تحديد مراكز الفئات للفئات المفتوحة . ولتلافي مثل هذه المشكلة قد نفترض حداً أدنى للفئات المفتوحة التي تقع في نهاية التوزيع بحيث يتناسب ذلك مع طبيعة البيانات موضع البحث . هذا وقد يتم أحياناً تحديد هذه الحدود اعتماداً على طول الفئة المعتمد في الفئات الأخرى ، مما يتيح لنا بالتالي اتباع الخطوات التي ورد ذكرها سابقاً في حساب المتوسط الحسابي . وتجدر الإشارة هنا أن المتوسط الحسابي الذي يتم الحصول عليه في مثل هذه الحالات هو تقريبي وليس دقيقاً . هناك بعض الباحثين يلجأ إلى

حذف الفئات المفتوحة إذا كان عدد تكرارتها قليلاً نسبياً إذا ما قورن بعدد التكرارات الكلية، وفي مثل هذه الحالة أيضاً يكون المتوسط الحسابي الناتج تقريبياً.

الخصائص العامة للمتوسط الحسابي:

هناك عدد من الخصائص التي يتميز بها المتوسط الحسابي يمكن اجمالها بما يلي:

1- يعد المتوسط الحسابي نقطة الاتزان للمشاهدات في توزيع معين، لأنه يمثل معدل تلك المشاهدات. فهو يشبه مركز الدائرة من حيث كونه النقطة التي تتجمع حولها معظم المشاهدات في التوزيع. فلو أخذنا مجموع انحرافات جميع المشاهدات في التوزيع عن وسطها الحسابي نجد أن قيمتها تساوي صفراً، ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:30): لديك المشاهدات التالية

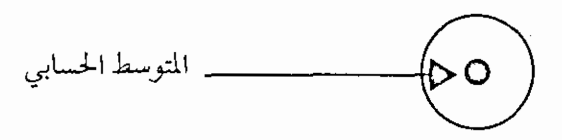
11, 9, 7, 8, 5

المتوسط الحسابي لها = $\frac{11+9+7+8+5}{5}$ =8

مجموع انحرافات القيم عن الوسط = ((5 - 8) + (8 - 8)+ (7 - 8) + (9 - 8) + (11 - 8))

= (3-) + صفر + (1-) + 1 + 3 = صفر

وهكذا نجد أن المتوسط الحسابي هو بمثابة نقطة التوازن للمشاهدات في التوزيع.



2- يعد المتوسط الحسابي من أكثر مقاييس النزعة المركزية استقراراً ولذلك فهو المفضل في حالة التحليلات الاحصائية المتقدمة.

3- تتأثر قيمة المتوسط الحسابي بالقيم المتطرفة في التوزيع سواء في بدايته أو نهايته بحيث تعمل هذه القيم المتطرفة على سحب المتوسط نحوها، وهذا بالتالي يجعل منه مؤشراً غير مناسب في حالة التوزيعات التكرارية الملتوية، ولتوضيح ذلك فلنأخذ الأمثلة التالية:

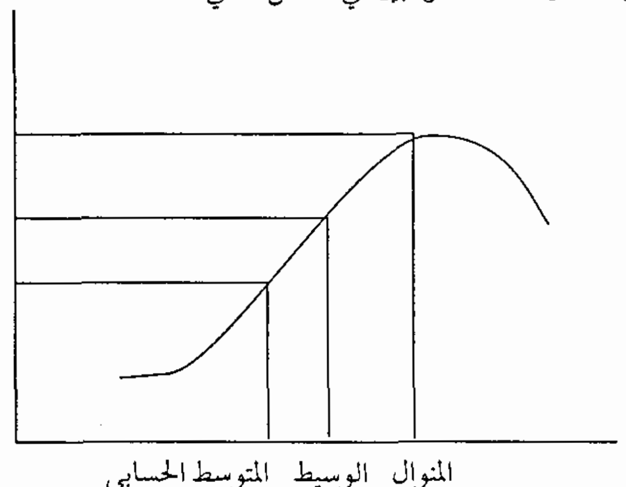
مثال (4:31) : لديك المشاهدات التالية :

63, 59, 58, 59, 56, 2, 2, 1

احسب المتوسط الحسابي لها ، ماذا تلاحظ ؟

$$\bar{س} = \frac{\sum س}{ن} = \frac{300}{8} = 37.5$$

نجد أن قيمة المتوسط الحسابي تأثرت بالقيم المتطرفة الموجودة في بداية التوزيع وهي (1 ,2 ,2) وبذلك نجد أن قيمته كانت منخفضة وغير متجانسة مع غالبية المشاهدات الأخرى في التوزيع . وهكذا نجد أن توزيع هذه البيانات يكون ملتوياً التواء سالباً (نحو اليسار) ، بحيث يكون المنوال أكبر من الوسيط والوسيط فيه أكبر من المتوسط ، وذلك كما هو مبين في الشكل التالي .



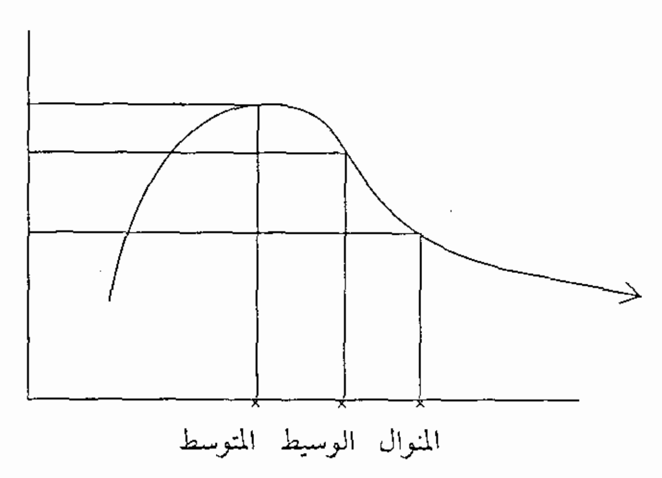
شكل (4:2): توزيع ملتو نحو اليسار (التوء سالباً)

وفيه المنوال > الوسيط > المتوسط

مثال (4:32) : لديك المشاهدات التالية : 2 ,1 ,3 ,5 ,4 ,95 ,110 احسب المتوسط الحسابي لها ، ماذا تلاحظ ؟

$$\bar{س} = \frac{\sum س}{(ن)} = \frac{210}{7} = 30$$

نلاحظ أن المتوسط تأثر بالقيم المتطرفة العالية في التوزيع وهي (95 , 100)، وبذلك نجد أن قيمته مرتفعة وغير متجانسة مع غالبية المشاهدات الأخرى في التوزيع. وهكذا نجد أن توزيع هذه البيانات هو ملتوٍ نحو اليمين (ملتوي التواءً موجباً) وفيه المتوسط أكبر من الوسيط والمنوال وذلك كما هو مبين بالشكل التالي :



شكل (4:3): توزيع ملتوي نحو اليمين (التواء موجب) وفيه الوسط > الوسيط > المنوال.

4- تتأثر قيمة المتوسط الحسابي بطول الفئة في حالة التوزيعات التكرارية المجدولة.

مثال (4:33): لديك توزيعان لمجموعة من المشاهدات طول الفئة في التوزيع الأول = 3، في حين طول الفئة = 7 في التوزيع الثاني، احسب المتوسط الحسابي في حالة كل توزيع. ماذا تلاحظ؟

الحل :

التوزيع الأول : طول الفئة = 3

الفئات	التكرار (ت)	م س	م س × ت
6 - 8	3	7	21
9 - 11	3	10	30
12 - 14	6	13	78
15 - 17	5	16	80
18 - 20	5	19	95
21 - 23	4	22	88
24 - 26	3	25	75
27 - 29	1	28	28
المجموع	30		495

$$\bar{س} = \frac{\sum م س \times ت}{\sum ت}$$

$$= \frac{495}{30} = 16.5$$

التوزيع الثاني : طول الفئة = 7

الفئات	التكرار (ت)	م س	م س × ت
7 - 13	10	10	100
14 - 20	12	17	204
21 - 27	8	24	192
المجموع	30		496

$$\bar{س} = \frac{\sum م س \times ت}{\sum ت}$$

$$= \frac{496}{30} = 16.53$$

وهكذا نلاحظ أن المتوسط الحسابي يتأثر بطول الفئة ، بحيث أنه يقترب تدريجياً من المتوسط الحقيقي للبيانات غير المجدولة مع زيادة طول الفئة .

5- تتأثر قيمة المتوسط الحسابي بالتحويلات الخطية التي تحدث على المشاهدات. حيث تزداد قيمته بمقدار القيمة الثابتة (س) التي تضاف لكل مشاهدة في التوزيع، كما أن قيمته تتناقص بنفس المقدار الثابت الذي يتم طرحه من قيم التوزيع. أما في حالة القسمة نجد أن قيمة المتوسط الحسابي تصبح عبارة عن ناتج قسمة هذا المتوسط على قيمة المقدار الثابت (س). وفي حالة ضرب القيم بمقدار ثابت (س)، فإن المتوسط الجديد يصبح مساوياً لقيمة المتوسط السابق مضروباً بقيمة المقدار الثابت (س). ولتوضيح ذلك فلنأخذ المثال التالي:

مثال (4:34): لديك المشاهدات التالية (5, 4, 3, 2, 1).

أ. احسب المتوسط الحسابي لهذه القيم

ب. ماذا يحدث للمتوسط الحسابي في حالة

1. إضافة مقدار ثابت (س) = 2 لكل مشاهدة.
2. طرح مقدار ثابت (س) = 2 لكل مشاهدة.
3. قسمة كل مشاهدة على مقدار ثابت (س) = 2.
4. ضرب كل مشاهدة بمقدار ثابت (س) = 2.

الحل:

أ. المتوسط = $\frac{\sum س}{ن} = \frac{5+4+3+2+1}{5} = \frac{15}{5} = 3$

ب. 1. عند إضافة مقدار ثابت (س) = 2 لكل مشاهدة، تصبح المشاهدات على النحو التالي: 3, 4, 5, 6, 7 ويصبح

المتوسط الحسابي: $\bar{س} = \frac{\sum س}{ن} = \frac{25}{5} = 5$ وهو عبارة عن المتوسط الحسابي السابق مضافاً إليه قيمة المقدار الثابت (س)؛ أي

$3 + 2 = 5$

2. عند طرح مقدار ثابت (س) = 2 من كل مشاهدة، تصبح المشاهدات على النحو التالي: -1، صفر, 1, 2, 3 ويصبح

المتوسط الحسابي الجديد: $\bar{س} = \frac{\sum س}{ن} = \frac{5}{5} = 1$ ، وهو عبارة عن المتوسط الحسابي السابق مطروحاً منه قيمة المقدار الثابت (س)؛ أي

$3 - 2 = 1$

3. عند قسمة كل مشاهدة على مقدار ثابت (س) = 2، تصبح المشاهدات على النحو التالي:
0.5, 1, 1.5, 2, 2.5 ويصبح

المتوسط الحسابي: $\bar{س} = \frac{\sum س}{ن} = \frac{7.5}{5} = 1.5$

وهو عبارة عن ناتج قسمة المتوسط الحسابي السابق على قيمة المقدار الثابت (س)؛ أي :

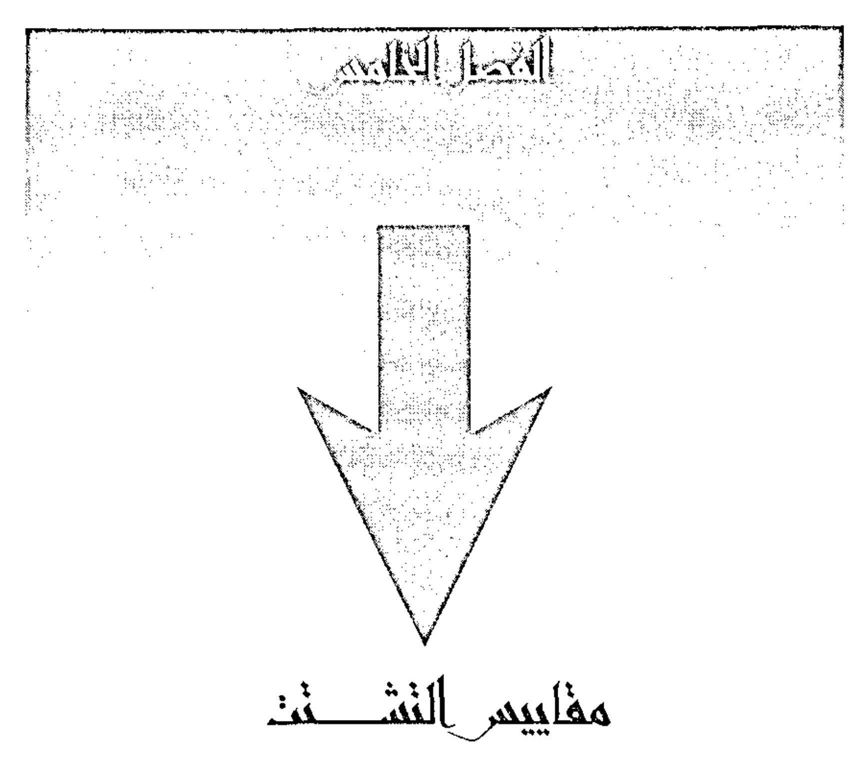
$3 \div 2 = 1.5$

4. عند ضرب كل مشاهدة بمقدار ثابت (س) = 2، تصبح المشاهدات على النحو التالي:
2, 4, 6, 8, 10 ويصبح

المتوسط الحسابي الجديد: $\bar{س} = \frac{\sum س}{ن} = \frac{30}{5} = 6$

وهو عبارة عن المتوسط الحسابي السابق مضروباً بقيمة المقدار الثابت (س)؛ أي:

$3 \times 2 = 6$.



مقاييس التشتت

Measures of Variation

تمهيد:

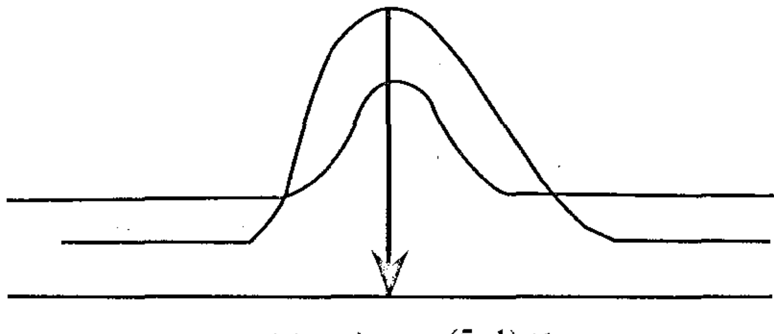
بالرغم من أهمية مقاييس النزعة المركزية التي ورد ذكرها سابقاً في تلخيص ووصف البيانات الاحصائية المأخوذة من مجتمع او عينة ما ، الا ان هذه المقاييس لا تعد كافية في اعطاء صورة واضحة عن توزيع البيانات الاحصائية . ولا سيما ان مثل هذه المقاييس لا تبيّن لنا مدى تباعد او تقارب المشاهدات (القيم) عن بعضها بعضاً وكيفية انتشارها . فلو اخذنا على سبيل المثال التوزيعين التاليين :

(140, 120, 105, 100, 95,)

(160, 125, 105, 95, 75)

نلاحظ ان المتوسط الحسابي والوسيط لهما متساويان وهما على التوالي (سَ=112 ، الوسيط=105) ، ولكن بالرغم من تماثل مقاييس النزعة المركزية لهذين التوزيعين فلا يمكن اعتبارهما متماثلين من حيث توزيع وانتشار البيانات ، فهي متجانسة ومتقاربة في التوزيع الاول ، الا انها متباعدة وغير متجانسة في التوزيع الثاني .

ولتوضيح الصورة بشكل افضل يمكن النظر الى الشكل التالي :



شكل (1 :5) يوضح مفهوم التشتت

في الشكل اعلاه نلاحظ توزيعيين تكراريين فيها تتشابه مراكز النزعة المركزية ولكنها تختلف من حيث انتشار المشاهدات (توزيع البيانات). وذلك لان مقاييس النزعة المركزية هي بمثابة مؤشرات مستوى او نقاط في توزيع معين ولا تعكس مدى تباعد او تباين المشاهدات. وانطلاقاً من ذلك لا بد من الاستعانة بمقاييس اخرى تسمى بمقاييس التشتت بالاضافة الى المقاييس السابقة من اجل اعطاء صورة كاملة عن طبيعة التوزيع ، اذ ان مثل هذه المقاييس تعد بمثابة مؤشرات تعبر عن المسافة اوالبعد بين المشاهدات في التوزيع.

واعتماداً على ذلك فإن مقاييس التشتت تستخدم كؤشرات لوصف مدى تقارب او تباعد القيم(المشاهدات) عن بعضها، بمعنى اذا كانت قيمة هذه المقاييس منخفضة ، فهذا يعني ان المشاهدات متجانسة وتقترب قيمها من بعضها، اما اذا كانت قيمتها مرتفعة فهذا دليل على عدم تجانس مثل هذه المشاهدات. هذا وتشتمل هذه المقاييس على مجموعة من المؤشرات تتمثل بما يلي :

1- المدى Range .

2- نصف المدى الربعي Interquaritle .

3- التباين Variance.

4- الانحراف المعياري Standard deviation.

أولاً: المدى Range

يرمز له بالرمز(م) وبالانجليزية بالرمز (R) وهو يعد اسهل طريقة لحساب التشتت مقارنة بمقاييس التشتت الاخرى واكثرها سهولة في الحساب، اذ يعتمد حسابه على تحديد الفرق بين اعلى وادنى قيمة في التوزيع، وذلك وفق المعادلة التالية :

المدى (م) = اعلى قيمة- ادنى قيمة.

مثال (5:1): لديك المشاهدات التالية وهي درجات مجموعة من الطلاب على امتحان اللغة العربية، احسب المدى لهذه الدرجات.

20, 19, 16, 10, 9, 16, 14, 17, 15, 8

المدى = اعلى مشاهدة- ادنى مشاهدة

= 20 - 8 = 12

هذا ويمكن تلخيص خصائص هذا المؤشر على النحو الآتي :

1- سهولة حسابه : حيث انه يمثل الفرق بين اعلى وادنى قيمة في التوزيع .

2- مؤشر سهل من حيث اعطاء صورة سريعة عن تشتت البيانات في توزيع معين .

3- لا تتأثر قيمته في حالة اجراء بعض التحويلات الخطية على البيانات مثل (الجمع والطرح) . فلو تم اضافة مقدار ثابت (س) لكل مشاهدة في التوزيع او تم طرح هذا المقدار من المشاهدات نجد ان قيمة المدى لا تتأثر بذلك .

لتوضيح ذلك فلنأخذ المثال التالي :

مثال(5:2) : التوزيع التالي يبين درجات (5) اشخاص على امتحان ما (20 ,24 ,26 ,28 ,30)

المطلوب :

أ- احسب المدى لهذه الدرجات؟

ب-اذا تم اضافة مقدار ثابت (س=2) لكل درجة في التوزيع ، فكم تصبح قيمة المدى .

جـ-اذا تم طرح مقدار ثابت (س=2) من كل درجة في التوزيع ، فكم تصبح قيمة المدى .

الحل:

أ- المدى = 30-20-10

ب-في حالة اضافة مقدار ثابت (س=2) لكل درجة في التوزيع ، عندها تصبح الدرجات الجديدة على النحو الآتي :

32, 30, 28, 26, 22

ويكون المدى لها : 32-22=10

جـ-في حالة طرح مقدار ثابت (س=2) من كل درجة في التوزيع ، عندها تصبح الدرجات الجديدة على النحو الآتي :

28, 26, 24, 22, 18

ويكون المدى لها : 28-18=10

وهكذا نلاحظ ان المدى لهذه الدرجات في الحالات الثلاث =10

حيث انه لم يتأثر بعمليات الجمع او الطرح التي اجريت على المشاهدات.

4- تتأثر قيمة المدى في بعض التحويلات الخطية (القسمة والضرب) التي تجرى على المشاهدات بحيث تتغير قيمته تبعاً لعمليات القسمة والضرب التي تجري على تلك المشاهدات.

مثال (5:3): لديك البيانات التالية:

18, 14, 12, 6, 4

المطلوب:

أ- احسب المدى لهذه البيانات؟

ب- في حالة قسمة كل مشاهدة في التوزيع على مقدار ثابت (س=2) ، فكم تصبح قيمة المدى؟

جـ-في حالة ضرب كل مشاهدة في التوزيع بمقدار ثابت (س=2) ، فكم تصبح قيمة المدى؟

الحل:

أ- المدى = 18-4=14

ب- في حالة قسمة كل مشاهدة في التوزيع على مقدار ثابت (س=2)، عندها تصبح المشاهدات الجديدة على النحو الآتي:

9, 7, 6, 3, 2

ويكون المدى لها 9-2 =(7)

جـ-في حالة ضرب كل مشاهدة بالتوزيع بمقدار ثابت (س=2) ، فإن القيم الجديدة تصبح على النحو الآتي:

36, 28, 24, 12, 8

ويكون المدى لها: 36-8=24

وهكذا نلاحظ ان قيمة المدى تتأثر بعمليات القسمة والضرب، حيث ان قيمته تتغير تبعاً لتغير قيمة المقدار الثابت التي تضرب به المشاهدات او تقسم عليه.

5- لا تدخل في حسابه كافة القيم الموجودة في التوزيع وانما يقتصر ذلك على قيمتين هما اقل واكبر

قيمة في التوزيع . وبذلك نجد ان قيمة المدى تتأثر بهاتين القيمتين من حيث كونهما متقاربتين او متباعدتين . وهذا يعني بالطبع ان قيمة المدى تتأثر بوجود القيم المتطرفة في التوزيع ، مما يجعل منه مؤشراً غير دقيق للحكم على مدى تجانس البيانات في حالة تلك التوزيعات .

ولتوضيح ذلك فلنأخذ المثال التالي :

مثال (5:4) : لديك التوزيع التالي :

30, 24, 18, 20, 22

المدى لهذا التوزيع : 30-18=12

ماذا يحدث للمدى اذا كانت المشاهدة الاخيرة 60 بدلاً من 30 .

المدى يصبح : 60-18=42

نلاحظ ان المدى تأثر بوجود القيمة المتطرفة ، مما جعل منه مؤشراً غير دقيق لوصف تشتت او تجانس البيانات .

6- تتأثر قيمة المدى بزيادة حجم العينة ، وذلك لأن زيادة حجم العينة من شأنه ان يزيد من احتمالية وجود قيم متطرفة .

7- يصلح المدى لاصدار الاحكام على بعض التوزيعات اذا تم استخدامه مع احد مقاييس النزعة المركزية مثل المتوسط . ولتوضيح ذلك فلنأخذ المثال التالي :

مثال (5:5) : تقدم خمسة اشخاص لثلاثة امتحانات على الرياضيات والفيزياء والكيمياء وكانت درجاتهم على التوالي كما هو مبين ادناه .

الشخص الأول : 45 ,85 ,95 : المتوسط الحسابي=75

الشخص الثاني : 75 ,70 ,80 : المتوسط الحسابي =75

الشخص الثالث : 60 ,90 ,75 : المتوسط الحسابي=75

الشخص الرابع : 60 ,90 ,95 : المتوسط الحسابي =75

الشخص الخامس : 60,65 ,95 : المتوسط الحسابي =75

وكما نلاحظ اعلاه انه بالرغم من ان المتوسطات الحسابية لدرجات هؤلاء الافراد الخمسة هي متساوية ، فلا يمكن القول انهم متماثلون من حيث مستوى الاداء على هذه المواد ، ولتحديد اي الطلاب

افضل يمكن استخدام المدى، حيث نلاحظ ان مدى الدرجات لهؤلاء الطلاب على المواد الثلاث كان على التوالي (50 للاول، 10 للثاني، 30 للثالث، 55 للرابع، 35 للخامس).

وعليه يمكن القول ان الطالب الثاني هوالافضل من حيث الاداء نظراً لان المدى بين درجاته هو الاقل وهذا يشير الى تجانس اداء الطالب على المواد الثلاثة مما يعكس اداءه الحقيقي.

ثانياً: المدى الربعي Interquaritle Range

يمثل المدى الربعي الفرق بين نقطتين في التوزيع هما المئين، 75 (الربع الثالث) والمئين 25 (الربع الأول). وبلغة اخرى يعرّف المدى الربعي على انه المسافة التي تقع بين المئين (ى) 75 والمئين (ى) 25. يشبه المدى الربعي المدى من حيث اعتماده على قيمتين فقط في حسابه ولكن بدلاً من الاعتماد على اقل وادنى قيمة كما هو الحال عند حساب المدى، نجد ان المدى الربعي يعتمد في حسابه على الربع الأول والربع الثالث في التوزيع وفق المعادلة التالية ك= ك3-ك 1 . وهذا يجعل منه اقل تأثراً بالقيم المتطرفة في التوزيع، وعليه ، فإن هذا المؤشر يستخدم لقياس التشتت في حالة التوزيعات التي تشتمل على قيم متطرفة او فئات مفتوحة.

يميل اغلب الاحصائيين الى استخدام نصف المدى الربعي في قياس درجة التشتت عوضاً عن المدى الربعي نفسه ويطلق عليه اسمه "Semi- quartile Range" وهوعبارة عن ناتج قسمة المدى الربعي على (2)، وفق المعادلة التالية: نصف المدى الربعي $= \frac{\text{ك } 3 - \text{ك}1}{2}$

ولحساب المدى الربعي يتم عادة اتباع نفس الخطوات المتبعة في حساب المئينات بحيث يتم ايجاد قيمة المئين (ى) 75 وقيمة المئين (ى)25، ثم يتم طرح قيمة المئين 25 من المئين 75 ليكون الناتج بمثابة المدى الربعي للتوزيع. اما نصف المدى الربعي فيتطلب حسابه قسمة المدى الربعي على العدد (2). ولتوضيح ذلك فلنأخذ الحالات المختلفة من التوزيعات.

اولاً: في حالة البيانات الخام:

يتم حساب كل من المدى الربعي ونصف المدى الربعي وفق الخطوات التالية :

1- تحديد قيمة الربع الاعلى (الثالث)، اي المئين 75 وقيمة الربع الادنى (الاول)، اي المئين 25.

2- ايجاد الفرق بين قيمتي الربعين (ي 75-ي 25) ، وبذلك نحصل على المدى الربعي.

3- نقسم المدى الربعي على العدد (2)، ونحصل على قيمة نصف المدى الربعي.

مثال (5:6) : لديك المشاهدات التالية : احسب كل من المدى الربعي ونصف المدى الربعي؟

24.22.21.16.13.12.9.7.6.4.3.2

الحل :

المئين 25 (الربع الاول) = $\frac{25}{100} \times 12 = (4)$؛ اي المشاهدة الرابعة في التوزيع وقيمتها (6) ، وعليه فإن قيمة المئين 25= الحد الفعلي للقيمة (6)= (6.5)

المئين 75 الربع الثالث = $\frac{75}{100} \times 12 = (9)$ ، اي المشاهدة التاسعة في التوزيع وقيمتها(16)، وعليه فإن قيمة المئين 75= الحد الفعلي للقيمة (16)= (16.5)

وعليه فإن المدى الربعي لهذا التوزيع= 16.5-6.5=10

اما نصف المدى الربعي فيساوي $\frac{10}{2} = 5$

ثانياً: في حالة البيانات غير المبوبة في جداول تكرارية:

لحساب كل من المدى الربعي ونصف المدى الربعي في مثل هذه الحالات يتم اتباع الخطوات التالية :

1- ترتيب المشاهدات تنازلياً أو تصاعدياً في عمود.

2- رصد التكرارات المقابلة لكل مشاهدة في عمود آخر.

3- تحديد التكرارات التراكمية لكل مشاهدة في عمود ثالث والحدود الفعلية في عمود رابع.

4- حساب التكرار التراكمي المقابل لكل من المئين 25 و المئين 75 .

5- ايجاد الفرق بين قيمتي المئينين 75و 25 ، ثم قسمة الناتج على العدد (2) وبذلك نحصل على نصف المدى الربعي.

مثال (5:7) :

لديك المشاهدات التالية :

8, 8, 8, 6, 6, 6, 5, 4, 4

20, 18, 18, 18, 18, 16, 16, 15, 12, 10, 8

احسب المدى الربعي ونصف المدى الربعي لهذه المشاهدات :

الحل:

نرتب البيانات ونحدد تكراراتها التراكمية كما هو مبين في الجدول ادناه:

المشاهدة	التكرار	التكرار التراكمي	الحدود العقلية
4	2	2	4.5
5	1	3	5.5
6	3	6	6.5
8	4	10	8.5
10	1	11	10.5
12	1	12	12.5
15	1	13	15.5
16	2	15	16.5
18	4	19	18.5
20	1	20	20.5
المجموع	20		

5

المئين 25

المئين 75

- نحسب المئين 25 على النحو الآتي :

التكرار التراكمي المقابل للمئين 25 $= \frac{25}{100} \times 20 = 5$

وبالنظر الى عمود التكرارات التراكمية نجد ان هذا التكرار يقع بين التكرارين التراكميين (3) و (6) اللذين يقابلان الحدين الفعليين (5.5) و (6.5) . وباستخدام النسبة والتناسب نجد قيمة هذا المئين على النحو الآتي :

3	تقابل	5.5
5 ←		→ س
6	تقابل	6,5

وعليه فإن 3 تقابل 1

2 تقابل س وبالضرب التبادلي نجد أن :

3 س = 2 ، س = $\frac{2}{3}$ = .67

وعليه فإن قيمة المئين 25 = 5.5 + .67- = 6.17

-نحسب المئين 75 على النحو الآتي :

التكرار التراكمي المقابل للمئين 75 = $\frac{75}{100}$ × 20 = 15

وبالنظر الى عمود التكرارات التراكمية في الجدول اعلاه نجد ان هذه القيمة موجودة وهي تقابل الحد الفعلي 16.5 وعليه فإن قيمة المئين 75 =16.5

اذاً المدى الربعي = المئين 75 - المئين 25

= 16.5 - 6.17 = 10.33 تقريباً

نصف المدى الربعي = $\frac{10.33}{2}$ = 5.165 تقريباً

ثالثاً : في حالة البيانات المبوبة في جداول تكرارية

لحساب كل من المدى ونصف المدى الربعي في حالة البيانات الجدولة في توزيعات تكرارية يتم اتباع نفس الخطوات الواردة سابقاً، ولكن يتم استخدام الحد الفعلي الاعلى للفئات عوضاً عن الحدود الفعلية للمشاهدات الحقيقية، وذلك كما هو موضح في المثال التالي :

مثال (5:8) : يبين الجدول التالي درجات مجموعة من الافراد على امتحان الكفاءة اللغوية، احسب نصف المدى الربعي لهذه الدرجات؟

الفئات	التكرار	التكرار التراكمي	الحدود الفعلية العليا
9-6	6	6	9.5
13-10	10	16	13.5
17-14	7	23	17.5
21-18	4	27	21.5
25-22	3	30	25.5
المجموع	30		

المئين 25 (7.5) ◄
المئين 75(22.5) ◄

الحل: نحسب المئين 25

-التكرار التراكمي المقابل للمئين 25 = $\frac{25}{100} \times 30 = 7.5$

- نلاحظ في عمود التكرارات التراكمية ان هذا التكرار يقع بين التكرارين التراكميين (6)و (16) والذين يقابلان الحدَّين الفعليين (9. 5) و (13. 5). وذلك على النحو التالي :

6	تقابل	9.5
7. 5 ◄──		──► س
16	تقابل	13. 5

وعليه فإن 10 تقابل 4

1.5 تقابل س

وبالضرب التبادلي (النسبة والتناسب) نجد ان 10س = 1.5 × 4

10س = 6

$س = \frac{6}{10} = 0.6$

وعليه فإن المئين 25 =9.5+0.6 =10.1

نحسب المئين 75

التكرار التراكمي للمئين 75 = $\frac{75}{100} \times 30 = 22.5$

وهذه القيمة تقع بين التكرارين التراكميين (16) و (23)

واللذين يقابلان الحدين الفعليين (13.5) و (17.5) وذلك على التالي :

	16	تقابل	13.5
	22.5 ⟵		⟶ س
	23	تقابل	17.5
اي ان	7	تقابل	4
	6.5	تقابل	س

وبالضرب التبادلي نجد ان 7س = 26

س = 3.71

وعليه فإن المئين 75 = 13.5+3.71 = 17.21 تقريباً

وهكذا فإن المدى الربعي=17.21-10.1=7.11

ويكون نصف المدى الربعي عبارة عن ناتج قسمة المدى الربعي على العدد (2)

اي 7.11 ÷ 2 = 3.56 تقريباً

ثالثاً : الانحراف المتوسط Mean Deviation

يعرف الانحراف المتوسط على أنه مطلق مجموع انحرافات القيم عن متوسطها الحسابي مقسوماً على عددها ، او مطلق مجموع انحرافات مراكز الفئات عن متوسطها الحسابي مقسوماً على مجموع التكرارات وفقاً للمعادلة التالية :

$$\text{الانحراف المتوسط (ن ح)} = \frac{|\text{س} - \bar{\text{س}}|}{\text{ن}}$$

فمن المعروف ان مجموع انحرافات اي قيم عن متوسطها الحسابي يساوي صفراً ، وذلك لان المتوسط الحسابي هو بمثابة نقطة الاتزان في اي توزيع ، وعليه ففي حساب الانحراف المتوسط يتم استخدام القيمة المطلقة للمجموع وذلك للتخلص من الاشارات السالبة التي تسبق بعض القيم وحتى لا يكون المجموع صفراً وذلك كما يظهر في المثال التالي :

مثال (5:9): امامك المشاهدات التالية 14 .16 .20 .22 .28 احسب الانحراف المتوسط لهذه المشاهدات:

الحل:

1- نحسب المتوسط الحسابي لها:

$$\bar{س} = \frac{\sum س}{ن} = \frac{100}{5} = 20$$

2- نحسب انحرافات القيم عن المتوسط الحسابي

وهي (14-20)، (16-20)، (20-20)، (22-20)، (28-20)

ويكون الناتج -6، -4، صفر، 2، 8

3- نستخدم القيمة المطلقة لهذه البيانات للتخلص من الاشارة السالبة.

بحيث تصبح الانحرافات على النحو التالي:

6، 4، صفر، 2، 8

4- نجد مجموع هذه الانحرافات المطلقة وذلك على النحو التالي:

6 + 4 + صفر +2 + 8 = 20

5- نقسم حاصل جمع الانحرافات على عددها ليكون الناتج هو الانحراف المتوسط للقيم:

$$\text{الانحراف المتوسط} = \frac{|س - \bar{س}|}{ن}$$

$$= \frac{20}{5} = 4$$

تجدر الاشارة هنا، ان مثل هذا المؤشر تندر استخداماته من قبل الاحصائيين، ولكن يمكن ان تكمن اهميته من حيث علاقته الوثيقة بمقاييس تشتت اخرى مثل الانحراف المعياري والتباين.

رابعاً: التباين The Variance

يعرف التباين على أنه متوسط مربع انحرافات القيم عن متوسطها الحسابي ، ويرمز له بالحرف ($ع^2$) باللغة العربية وبالرمز اللاتيني (σ^2) او الحرف (S^2) باللغة الانجليزية .

يعد التباين من اكثر المقاييس استخداماً ولا سيما في عمليات الاحصاء التحليلي ، هذا وتتأثر قيمته بالقيم المتطرفة بحيث لا ينصح باستخدامه في حالة وجود قيم متطرفة في توزيع معين .

وبشكل عام يتم حساب التباين عبر الخطوات التالية :

1- حساب المتوسط الحسابي للقيم في التوزيع .

2- حساب انحرافات القيم عن المتوسط الحسابي .

3- تربيع الانحرافات

4- قسمة مربع الانحرافات على مجموع المشاهدات (القيم) بحيث يشكل الناتج التباين للمشاهدات . وبالرموز يمكن التعبير عن معادلة ايجاد التباين بالمعادلة التالية :

$$ع^2 = \frac{\sum (س-\bar{س})^2}{ن}$$

وتجدر الاشارة هنا ، انه في حالة كون البيانات تمثل مجتمعاً احصائياً معيناً فعندها يتم قسمة مجموع مربع الانحرافات على عدد الافراد في ذلك المجتمع (ن)، اما في حالة كون البيانات مأخوذة من عينة او مجموعة جزئية من مجتمع احصائي ، فعندها يتم قسمة ناتج مجموع مربعات الانحرافات على (ن- 1)، اي عدد افراد العينة مطروحاً منها العدد (1) .

مثال (5:10): لديك المشاهدات التالية وهي تمثل درجات عدد من الافراد على امتحان تحصيلي .

10. 9. 11. 13. 15. 12. 10. 7. 8. 5

احسب التباين لهذه القيم في حالة :

أ- كونها تمثل مجتمعا احصائيا (جميع الحالات في المجتمع) .

ب- كونها تمثل عينة جزئية من المجتمع الاحصائي .

الحل:

1- نحسب المتوسط الحسابي لهذه الدرجات

$$\bar{س} = \frac{\sum س}{ن} = \frac{100}{10} = 10$$

2- نحسب انحرافات القيم عن المتوسط الحسابي وذلك على النحو التالي:

(10-5)،(10-8)، (10-10)،(10-7)،(10-12)،(10-15)، (10-13)،(10-11)، (10-10)،(10-9).

ويكون الناتج .

-5، -2، صفر، -3، 2، 5، 3، 1، -1، صفر

3- نحسب مربع الانحرافات عن المتوسط ثم نجد مجموعها:

25 + 4 + صفر + 9 + 4+ 25 + 9 + 1+ 1 +صفر = 78

4- ولتحديد قيمة التباين نقسم مجموع مربع الانحرافات على (ن) في حالة أنها تمثل المجتمع الكلي وعلى (ن -1) في حالة كونها تمثل عينة من ذلك المجتمع . وعليه تكون اجابة الفرع (أ) على النحو الآتي:

$$\text{التباين } (ع^2) = \frac{\sum (س - \bar{س})^2}{ن}$$

$$\frac{78}{10} = 7.8$$

اما اجابة الفرع (ب) تكون على النحو الآتي:

$$\text{التباين } (ع^2) = \frac{\sum (س - \bar{س})^2}{ن - 1}$$

$$\frac{78}{9} = 8.67$$

حساب التباين في حالة البيانات غير المبوبة في جداول تكرارية.

يتم حساب التباين في حالة البيانات غير الجدوله وفق المعادلة التالية :

$$\text{التباين } (ع^2) = \frac{\sum (س - \bar{س})^2 \times ت}{\sum ت}$$

، وذلك وفق الخطوات التالية :

1- حساب المتوسط الحسابي لهذه المشاهدات .

2- ايجاد انحرافات المشاهدات عن المتوسط الحسابي .

3- ايجاد ناتج ضرب الانحرافات في التكرارات المقابلة لها .

4- ايجاد مربع حاصل ضرب الانحرافات في التكرارات .

5- قسمة الناتج على (ن أو $\sum$ ت) في حالة بيانات المجتمع وعلى (ن-1) أو ($\sum$ ت-1) في حالة بيانات العينة للحصول على التباين .

مثال (5:11): الجدول التالي يلخص اوزان مجموعه من الطلاب ، احسب التباين لهذه الاوزان .

المشاهدات (الاوزان)	التكرار	س×ت	($س-\bar{س}$)	$(س-\bar{س})^2$	$(س-\bar{س})^2$×ت
24	2	48	6.34-	40.20	80.40
25	3	75	5.34-	28.50	85.50
27	4	108	3.34-	11.12	44.48
30	7	210	.34-	0.12	0.48
32	5	160	1.66	2.76	13.80
34	6	204	3.66	13.4	80.40
35	3	105	4.66	21.72	65.16
المجموع	30	910			370.58

الحل :

(1) نحسب المتوسط الحسابي لهذه الاوزان :

$$\bar{س} = \frac{\sum س \times ت}{\sum ت} = \frac{910}{30} = 30.34$$

(العمود الثالث)

(2) نحسب انحرافات القيم عن المتوسط الحسابي (العمود الرابع)

(3) نحسب مربع انحرافات القيم عن المتوسط الحسابي (العمود الخامس)

(4) نحسب ناتج ضرب مربع الانحرافات في التكرارات ، ثم نحسب المجموع الكلي (العمود السادس).

(5) نقسم الناتج في الخطوة رقم (4) على مجموع التكرارات ($\sum$ ت) لنحصل على التباين في حالة بيانات المجتمع وعلى ($\sum$ ت-1) في حالة بيانات العينة. وذلك على النحو الاتي:

$$\text{التباين للمجتمع} = \frac{\sum (\text{س} - \bar{\text{س}})^2 \times \text{ت}}{\sum \text{ت}}$$

$$= \frac{370.58}{30} = 12.35 \text{ تقريبا}$$

$$\text{التباين للعينة} = \frac{\sum (\text{س} - \bar{\text{س}})^2 \times \text{ت}}{\sum \text{ت} - 1}$$

$$= \frac{370.58}{29} = 12.78 \text{ تقريبا}$$

حساب التباين في حالة البيانات المبوبة في توزيعات تكرارية

يتم حساب التباين في حالة البيانات المبوبة في توزيعات تكرارية وفق المعادلة التالية:

$$\text{ع}^2 = \frac{\sum (\text{م س} - \bar{\text{س}})^2 \times \text{ت}}{\sum \text{ت}}$$ ، حيث م س : مركز الفئة.

تتمثل خطوات حسابه بالآتي:

أولاً: يتم تحديد مراكز الفئات.

ثانياً: يتم حساب المتوسط الحسابي.

ثالثاً: يتم تحديد انحرافات مراكز الفئات عن المتوسط الحسابي.

رابعاً: يتم حساب مربع انحرافات مراكز الفئات عن المتوسط الحسابي.

خامساً: يتم ايجاد حاصل ضرب مربع الانحرافات في التكرارات، ثم تحديد المجموع الكلي.

سادساً: يتم ايجاد ناتج قسمة المجموع الكلي على مجموع التكرارات ($\sum$ ت) في حالة بيانات المجتمع وعلى ($\sum$ ت-1) في حالة بيانات العينة، بحيث يشكل هذا الناتج التباين للتوزيع.

مثال (5:12) : الجدول التالي يلخص درجات (100) فرد على اختبار القدرة الرياضية ، احسب التباين لهذه الدرجات :

الفئات	التكرار	م س	ت ×م س	(م س – سَ)	^{2}(م س–سَ)	(م س – سَ)^2× ت
24-20	3	22	066	20.80-	432.64	1297.92
29-25	7	27	189	15.80-	249.64	1747.48
34-30	10	32	320	10.80-	116.64	1166.40
39-35	13	37	481	5.80-	33.64	437.32
44-40	22	42	924	0.80-	0.64	14.08
49-45	18	47	846	4.20	17.64	317.52
54-50	17	52	884	9.20	84.64	1438.88
59-55	10	57	570	14.20	201.64	2016.40
المجموع	100		4280			8436

الحل :

1- يتم تحديد مراكز الفئات (العمود الثالث في الجدول)

2- حساب المتوسط الحسابي سَ = $\frac{\sum(\text{م س} \times \text{ت})}{\sum \text{ت}} = \frac{4280}{100} = 42.8$

3- ايجاد انحرافات مراكز الفئات عن المتوسط الحسابي (العمود الخامس).

4- ايجاد مربع انحرافات مراكز الفئات عن المتوسط الحسابي (العمود السادس).

5- ايجاد حاصل ضرب مربع الانحرافات في التكرار (العمود السابع).

6- نقسم الناتج في خطوة (5) على مجموع التكرارات ($\sum$ ت) .

في حالة بيانات المجتمع او على ($\sum$ ت-1) في حالة بيانات العينة للحصول على التباين.

التباين (ع^2) = $\frac{\sum(\text{م س} - \text{سَ})^2 \times \text{ت}}{\sum \text{ت}}$ في حالة بيانات المجتمع .

$= \frac{8436}{100} = 84.36$

التباين $(ع^2) = \dfrac{\sum (\text{م س} - \bar{\text{س}})^2 \times \text{ت}}{\sum \text{ت} - 1}$ في حالة بيانات العينة .

$$= \frac{8436}{99} = 85$$

طريقة اخرى لحساب التباين:

هناك طريقة اخرى لحساب التباين تعرف بالطريقة العامة ، ومثل هذه الطريقة يفضل استخدامها في حالة كون المتوسط الحسابي يشتمل كسراً عشرياً ، حيث اننا في هذه الطريقة لا نلجأ الى ايجاد الانحرافات عن المتوسط ، وانما يستعاض عن ذلك بالتعامل مع نفس القيم وفقاً للمعادلة التالية :

في حالة بيانات المجتمع : التباين = $(ع^2) = \dfrac{\sum \text{س}^2 - \dfrac{(\sum \text{س})^2}{\text{ن}}}{\text{ن}}$

في حالة بيانات العينة : التباين = $(ع^2) = \dfrac{\sum \text{س}^2 - \dfrac{(\sum \text{س})^2}{\text{ن}}}{\text{ن} - 1}$

ومثل هذه المعادلة تستخدم في حالة البيانات الخام .

هذا وتكون المعادلة السابقة في حالة البيانات غير المبوبة على النحو التالي :

التباين لبيانات المجتمع : $(ع^2) = \dfrac{\sum \text{س}^2\text{ت} - \dfrac{(\sum \text{س ت})^2}{\sum \text{ت}}}{\sum \text{ت}}$

التباين لبيانات العينة = $(ع^2) = \dfrac{\sum \text{س}^2\text{ت} - \dfrac{(\sum \text{س ت})^2}{\sum \text{ت}}}{\sum \text{ت} - 1}$

اما في حالة البيانات المبوبة في جداول تكرارية فتكون معادلة التباين على النحو الآتي :

$$\text{للمجتمع :}\quad ع^2 = \frac{\sum \text{م س}^2\,ت - \frac{\left(\sum \text{م س ت}\right)^2}{\sum ت}}{\sum ت}$$

$$\text{للعينة :}\quad ع^2 = \frac{\sum \text{م س}^2\,ت - \frac{\left(\sum \text{م س ت}\right)^2}{\sum ت}}{\sum ت - 1}$$

مثال (5:13) : احسب التباين للمشاهدات التالية :

13, 5, 14, 12, 6

الحل:

1- نحسب مجموع مربعات القيم $\left(\sum س^2\right)$ على النحو الآتي :

$(13)^2 + (5)^2 + (14)^2 + (12)^2 + (6)^2$

570 = 169 + 25 + 196 + 144 + 36 =

2- نحسب مربع مجموع القيم $\left(\sum س\right)^2$ على النحو الآتي

$2500 = (50)^2 = (13+5+14+12+6)^2$

3 - نعوض هذه القيم في المعادلة التالية :

للمجتمع : $ع^2 = \dfrac{\sum س^2 - \dfrac{\left(\sum س\right)^2}{ن}}{ن}$

$$= \frac{570 - \frac{2500}{5}}{5}$$

$$= \frac{500\text{-}570}{5} = \frac{70}{5} = 14$$

للعينة : $ع^2 = \dfrac{\sum س^2 - \dfrac{\left(\sum س\right)^2}{ن}}{ن - 1}$

$$= \frac{570 - \frac{2500}{5}}{5 - 1}$$

$$= \frac{570\text{-}500}{4} = \frac{70}{4} = 17.5$$

مثال (5:14) : احسب التباين للبيانات في الجدول ادناه :

الوزن (س)	التكرار	س×ت	$س^2$	$س^2$×ت
24	2	48	576	1152
25	3	75	652	1875
27	4	108	729	2916
30	7	210	900	6300
32	5	160	1024	5120
34	6	204	1156	6936
35	3	105	1225	3675
المجموع	30	910		27974

الحل:

1- نحسب ناتج ضرب المشاهدات في التكرارات المقابلة لها (العمود الثالث).

2- نحسب مربعات قيم س (العمود الرابع).

3- نجد ناتج حاصل ضرب مربعات القيم في التكرارات المقابلة لها(العمود الخامس).

4- لتحديد قيمة التباين يتم تعويض القيم السابقة في المعادلة التالية:

للمجتمع : التباين $\text{ع}^2 = \dfrac{\sum \text{س}^2 \times \text{ت} - \dfrac{\left(\sum \text{س} \times \text{ت}\right)^2}{\sum \text{ت}}}{\sum \text{ت}}$

$$= \frac{27974 - \frac{(919)^2}{30}}{30}$$

$$= \frac{27974 - \frac{828100}{30}}{30}$$

$$= \frac{27974 - 27603.34}{30} = \frac{370.66}{30} = 12.35$$

التباين للعينة= $\text{ع}^2 = \dfrac{\sum \text{س}^2 \times \text{ت} - \dfrac{\left(\sum \text{س}\,\text{ت}\right)^2}{\sum \text{ت}}}{\sum \text{ت} - 1}$

$$\text{ع}^2 = \frac{27974 - 27603.34}{29}$$

$= \dfrac{370.66}{29} = 12.78$ تقريبا

مثال (5:15) احسب التباين للبيانات في الجدول التالي :

الفئات	التكرار	م س	م س× ت	م س^2	م س^2 × ت
24-20	3	22	66	484	1452
29-25	7	27	189	729	5103
34-30	10	32	320	1024	10240
39-35	13	37	481	1369	17797
44-40	22	42	924	1764	38808
49-45	18	47	846	2209	39762
54-50	17	52	884	2704	45968
59-55	10	57	570	3249	32490
المجموع	100		4280		191620

الحل :

(1) نحسب مراكز الفئات (م س) (العمود الثالث)

(2) نجد ناتج ضرب مراكز الفئات في التكرار (م س×ت) العمود الرابع

(3) نحسب مربعات مراكز الفئات (م س2) العمود الخامس

(4) نحسب مجموع ناتج ضرب مربعات مراكز الفئات في التكرار (م س2× ت) العمود السادس.

(5) نعوض القيم في المعادلة التالية لتحديد قيم التباين

في حالة بيانات المجتمع :
$$ع^2 = \frac{\sum م س^2 \times ت - \frac{\left(\sum م س \times ت\right)^2}{\sum ت}}{\sum ت}$$

$$= \frac{191620 - \frac{(4280)^2}{100}}{100}$$

$$= \frac{\frac{18378400}{100} - 191620}{100}$$

$$= \frac{183184-191620}{100} = \frac{8436}{100} = 84.36$$

التباين في حالة بيانات العينة : $ع^2 = \dfrac{\sum \text{م س}^2 \times \text{ت} - \dfrac{\left(\sum \text{م س} \times \text{ت}\right)^2}{\sum \text{ت}}}{\sum \text{ت} - 1}$

$$= \frac{183184 - 191620}{99}$$

$$= \frac{8436}{99} = 85.29$$

خامساً : الانحراف المعياري Standard deviation

يعد الانحراف المعياري اكثر مقاييس التشتت انتشاراً وشيوعاً ، فهو يحتل اهمية بالغة من حيث استخداماته المتعددة في العديد من التحليلات الاحصائية المتقدمة مثل اختبار الفرضيات ، وايجاد فترات الثقة ، واجراء المقارنات الاحصائية وتقدير العلامات المعيارية والى غير ذلك من العمليات الاخرى .

يرمز للانحراف المعياري بالحرف العربي (ع) وبالرمز اللاتيني (σ) وهو عبارة عن الجذر التربيعي لمتوسط مجموع انحرافات القيم عن متوسطها الحسابي . وبلغه اخرى فهو يمثل الجذر التربيعي للتباين .

يتطلب حساب الانحراف المعياري عمليات احصائية اكثر تعقيداً من تلك التي تستخدم في حساب المدى او المدى الربعي ، وعموماً تتطلب عملية حسابه اتباع نفس الخطوات التي تستخدم لحساب الانحراف المعياري ، بحيث يصار الى اخذ الجذر التربيعي للتباين . وبالرموز فان معادلة حساب التباين هي .

$$ع = \sqrt{\frac{\sum (\text{س} - \bar{\text{س}})^2}{\text{ن}}} \quad \text{أو} \quad ع = \sqrt{\frac{\sum \text{س}^2 - \dfrac{(\sum \text{س})^2}{\text{ن}}}{\text{ن}}}$$

تستخدم المعادلة السابقة في حالة البيانات او المشاهدات الخام، اما في حالة البيانات غير المبوبة في توزيعات تكرارية نستخدم المعادلة التالية :

$$ع = \sqrt{\frac{\sum ت(س-\bar{س})^2}{\sum ت}}$$

أو المعادلة التالية :

$$ع = \sqrt{\frac{\sum س^2 \times ت - \frac{(\sum س \times ت)^2}{\sum ت}}{\sum ت}}$$

وفي حالة البيانات المبوبة في توزيعات تكرارية ، يتم استخدام المعادلة التالية لحساب الانحراف المعياري

$$ع = \sqrt{\frac{\sum ت (\sum م س - \bar{س})^2}{\sum ت}}$$

أو المعادلة التالية :

$$ع = \sqrt{\frac{\sum م س^2 \times ت - \frac{(\sum م س - ت)^2}{\sum ت}}{\sum ت}}$$

ومن الجدير ذكره ان المعادلات السابقة تستخدم في حالة بيانات احصائية تتعلق بالمجتمع ، اما في حالة البيانات المأخوذة من عينة ، فيتم القسمة على (ن-1) أو $\sum$ت -1وذلك وفق المعادلات التالية :

(1) في حالة البيانات الخام

$$ع = \sqrt{\frac{\sum (س-\bar{س})^2}{ن - 1}}$$

أو

$$\sqrt{\frac{\sum س^2 - \frac{(\sum س)^2}{ن}}{ن - 1}}$$

(2) في حالة البيانات غير المبوبة في توزيعات تكرارية

$$\text{ع} = \sqrt{\frac{\sum \text{ت}(\text{س}-\bar{\text{س}})^2}{\sum \text{ت}}}$$

أو المعادلة التالية :

$$\text{ع} = \sqrt{\frac{\sum \text{س}^2 \times \text{ت} - \frac{(\sum \text{س} \times \text{ت})^2}{\sum \text{ت}}}{\sum \text{ت} - 1}}$$

(3) في حالة البيانات المبوبة في توزيعات تكرارية

$$\text{ع} = \sqrt{\frac{\sum (\text{م س} - \bar{\text{س}})^2 \times \text{ت}}{\sum \text{ث} - 1}}$$

أو

$$\text{ع} = \sqrt{\frac{\sum \text{م س}^2 \times \text{ت} - \frac{(\sum \text{م س} \times \text{ت})^2}{\sum \text{ت}}}{\sum \text{ت} - 1}}$$

مثال (5:16) : احسب الانحراف المعياري للبيانات المبينة في الجدول ادناه

س	ت	س×ت	(س − سَ)	$(\text{س}-\bar{\text{س}})^2$	$(\text{س}-\text{س})^2 \times \text{ت}$
15	3	45	3.95-	15.60	46.81
17	6	102	1.95-	3.80	22.82
18	7	126	.95-	.90	6.32
19	12	228	0.05	.025	.03
20	8	160	1.05	1.10	8.82
24	3	72	5.05	25.50	76.51
25	1	25	6.05	36.60	36.60
المجموع	40	758			197.91

الحل:

أولاً: باستخدام معادلة الانحرافات

$$ع = \sqrt{\frac{\sum (س - سَ)^2 \times ت}{\sum ت - 1}}$$

نستخدم نفس الخطوات المتبعة في حساب التباين وذلك كما هو مبين في الجدول اعلاه.

المتوسط الحسابي = $\frac{758}{40} = 18.95$

ثم نسحب مربع انحراف كل قيمة عن المتوسط الحسابي ونضرب الناتج بالتكرار المقابل لها و نعمل على ايجاد المجموع الكلي، بحيث يصار على قسمته على مجموع التكرارت، ثم اخذ الجذر التربيعي للناتج ليمثل الانحراف المعياري. ومثل هذه الخطوات موضحة بالجدول السابق.

الانحراف المعياري (ع)

$$(ع) = \sqrt{\frac{197.91}{40-1}}$$

$$= \sqrt{\frac{197.91}{39}}$$

$$\sqrt{5.075} = 2.25$$

ثانياً: باستخدام المعادلة العامة

$$ع = \sqrt{\frac{\sum س^2 \times ت - \frac{(\sum س \times ت)^2}{\sum ت}}{ت - 1}}$$

كذلك نستخدم نفس الخطوات المتبعة في حساب التباين، وذلك كما هو مبين في الجدول التالي:

س	ت	س×ت	$س^2$	$س^2$×ت
15	3	45	225	675
17	6	102	289	1734
18	7	126	324	2268
19	12	228	361	4332
20	8	160	400	3200
24	3	072	576	1728
25	1	025	625	625
المجموع	40	758		14562

$$ع = \sqrt{\frac{\sum س^2 \times ت - \frac{(\sum س \times ت)^2}{\sum ت}}{ت-1}}$$

$$= \sqrt{\frac{14562 = \frac{(758)^2}{40}}{40-1}}$$

$$= \frac{\sqrt{14562 - \frac{574564}{40}}}{39}$$

$$= \sqrt{\frac{14562 - 14364.1}{39}}$$

$$= \sqrt{\frac{197.9}{39}} = \sqrt{5.75}$$

= 2.25 تقريباً

الخصائص العامة للتباين والانحراف المعياري:

يتميز كل من التباين والانحراف المعياري بالمميزات التالية :

اولاً : كلاهما يعد مقياس بعد او مسافة وليس مؤشر نقطة او مستوى ، ولذلك نجد ان قيمهما دائما موجبة ولا يمكن ان تكون سالبة .

ثانياً: كلاهما من اكثر مقاييس التشتت شيوعاً واستخداماً ولا سيما في العمليات الاحصائية المتقدمة مثل تحليل التباين ، اختبار الفرضيات ، تقدير العلامات المعيارية ، ايجاد فترات الثقة واجراء المقارنات الاحصائية ، ويكاد يكون الانحراف المعياري اكثر اهمية من التباين وذلك لان التباين لا يقاس بنفس وحدات المتغير وانما بوحدات مربعة . في حين نجد ان الانحراف المعياري يقاس بنفس وحدات المتغير موضع البحث .

ثالثاً:تتأثر قيمتهما بوجود قيم متطرفه في التوزيع . والسبب في ذلك ان حسابهما يعتمد على المتوسط الحسابي . ونظراً لتأثر قيمة المتوسط الحسابي بوجود القيم المتطرفة في التوزيع سواء المتدنية او العالية منها . فإن انحرافات القيم عن ذلك المتوسط سوف تتأثر بالتالي ، وهذا ينعكس في قيمة التباين والانحراف المعياري . وعليه لا يفضل استخدام كل من التباين والانحراف المعياري في حالة التوزيعات التي تشتمل على قيم متطرفة .

رابعاً: لا تتأثر قيمتهما بالتحويلات الخطية من جمع وطرح والتي تجري على المشاهدات في التوزيع . والسبب في ذلك ان المتوسط الحسابي يزداد او ينقص بنفس قيمة المقدار الذي يضاف او يطرح من المشاهدات ، حيث نجد أن قيمة المتوسط الحسابي تتغير بنفس قيمة المقدار الثابت في حالة الطرح والجمع . وعليه فإن انحرافات القيم عند ذلك المتوسط تبقى ثابتة ، وهذا بالتالي لا يؤثر في قيمة كل من التباين والانحراف المعياري . اما في حالة الضرب والقسمة نجد ان قيمتهما تتغير وذلك لان الانحرافات عن المتوسط تتغير تبعاً لعمليات الضرب والقسمة ، وهذا بالتالي يؤدي الى تغيير قيمتهما ولتوضيح ذلك فلنأخذ المثال التالي :

مثال (5:17) : لديك المشاهدات التالية

5.8.7.6.4

المطلوب :

أ- احسب كل من المتوسط والتباين والانحراف المعياري للمشاهدات السابقة :

ب- اذا تم اضافة مقدار ثابت قيمته $\sum$(س) = 2 لكل مشاهدة في التوزيع ، ماذا يحدث لكل من المتوسط الحسابي والتباين والانحراف المعياري .

جـ- اذا تم ضرب كل مشاهدة في التوزيع بمقدار ثابت قيمته س=2 ، ماذا يحدث لكل من المتوسط الحسابي والتباين والانحراف المعياري ؟

الحل:

أ- المتوسط الحسابي (سَ) = $\frac{\sum س}{ن}$

$= \frac{30}{5} = 6$

التباين $ع^2 = \frac{\sum (س-سَ)^2}{ن}$

$ع^2 = \frac{(6-5)^2+(6-8)^2+(6-7)^2+(6-6)^n+(6-4)^2}{5}$

$= \frac{4 + \text{صفر} + 1 + 4 + 1}{5} = \frac{10}{5}$

$= ع^2 = 2$

الانحراف المعياري $ع = \sqrt{2} = \sqrt{2} = 1.41$

ب)في حالة اضافة مقدار ثابت (س=2) لكل مشاهدة في التوزيع ، فإن المشاهدات الجديدة تصبح على الآتي :

7. 10. 9. 8. 6

عليه فإن المتوسط الحسابي (سَ) = $\frac{\sum س}{ن} = \frac{40}{5} = 8$ ، لاحظ أن قيمة المتوسط ازدادت بنفس قيمة المقدار الثابت الذي تمت اضافته لكل المشاهدات في التوزيع .

التباين $ع^2 = \frac{\sum (س-سَ)^2}{ن}$

$= \frac{(8-7)^2+(8-10)^2+(8-9)^2+(8-8)+(8-6)^2}{5} = \frac{10}{5} = 5$

الانحراف المعياري $ع = \sqrt{2} = \sqrt{2} = 1.41$

لاحظ هنا، ان المتوسط الحسابي ازداد بنفس قيمة المقدار الثابت الذي تمت اضافته لجميع المشاهدات، وكنتيجة لذلك فإن انحرافات القيم عن متوسطها الحسابي لم تتغير، بل بقيت ثابته وهذا بالتالي لم يؤثر في قيمة كل من التباين والانحراف المعياري.

جـ) في حالة ضرب المشاهدات في مقدار ثابت(س = 2)، فإن قيم التوزيع الجديدة تصبح على النحو الآتي:

10. 16. 14. 12. 8

المتوسط الحسابي: $\bar{س} = \frac{\sum س}{ن} = \frac{60}{5} = 12$

لاحظ ان قيمة المتوسط الجديد هي ناتج ضرب المتوسط السابق في المقدار الثابت الذي تمت اضافته لجميع المشاهدات في التوزيع.

التباين $ع^2 = \frac{\sum (س-\bar{س})^2}{ن}$

$$= \frac{(12-8)^2 + (12-12)^2 + (12-14)^2 + (12-16)^2 + (12-10)^2}{5}$$

$$= \frac{16 + \text{صفر} + 4 + 16 + 4}{5} = \frac{40}{5} = 8$$

لاحظ أن قيمة التباين تضاعفت في حالة الضرب.

الانحراف المعياري ع = $\sqrt{8}$

= $\sqrt{8}$ = 2.83 تقريباً

وكما هو الحال في عمليات الجمع، فإن قيمة كل من التباين والانحراف المعياري لا تتأثر في حالة الطرح، الا انها تتغير في عمليات القسمة كما هو الحال في عمليات الضرب، جرب ذلك.

خامساً: تدخل في حسابهما كافة القيم التي يشتمل عليها التوزيع، وهذا يجعل منهما أكثر مقاييس التشتت استقراراً ولا سيما في حالة التوزيعات التي لا تشتمل قيماً متطرفة.

سادساً: تتأثر قيمتهما بطول الفئة في حالة البيانات المبوبة في توزيعات تكرارية. والسبب في ذلك ان قيمة المتوسط الحسابي تتأثر أيضاً بطول الفئة. وكنتيجة لذلك فإن انحرافات مراكز الفئات على

المتوسط تتأثر وهذا بالتالي يؤثر في قيمة كل منهما . بإمكانك التأكد من ذلك بحساب كل من التباين والانحراف المعياري لتوزيع من المشاهدات بطول الفئة (4) وايضاً بطول فئة (8)ولاحظ الفرق .

التباين العام (الموزون)

في بعض الحالات تتوفر لدينا مؤشرات احصائية مثل المتوسط الحسابي والانحراف المعياري لمجموعات جزئية على متغير ما في الوقت الذي يصعب الحصول على مثل هذه المؤشرات مباشرة للعينة الكلية أو المجتمع . ففي مثل هذه الحالات ، يتم تقدير التباين والانحراف المعياري للمجتمع او العينة الكلية ، وهو (ما يسمى بالتباين او الانحراف المعياري العام) من خلال استخدام المعلومات الجزئية المتوفرة لدينا ، وذلك وفق المعادلة التالية :

$$ع^2 = \frac{(ن_1-1)\times ع_1^2 + (ن_2-1)\times ع_2^2 + (ن_3-1)\times ع_3^2 + ن_1(\bar{س}_1-\bar{س})^2 + ن_2(\bar{س}_2-\bar{س})^2 + ن_3(\bar{س}_3-\bar{س})^2 \ldots}{ن_1 + ن_2 + ن_3 - 1}$$

$ن_1$: هوعدد الافراد في المجموعة الاولى .

$\bar{س}$: المتوسط الحسابي لافراد المجموعة الاولى .

$ن_2$: عدد الافراد في المجموعة الثانية .

$\bar{س}_2$: المتوسط الحسابي لافراد المجموعة الثانية .

$ن_3$: عدد الافراد في المجموعة الثانية .

$\bar{س}_3$: المتوسط الحسابي لافراد المجموعة الثالثة .

$\bar{س}$: هو الوسط الموزون ويساوي :

$$\bar{س} = \frac{\bar{س}_1 \times ن_1 + \bar{س}_2 \times ن_2 + \bar{س}_3 \times ن_3}{ن_1 + ن_2 + ن_3}$$

مثال (5:18): فيما يلي مؤشرات احصائية حول اداء افراد ثلاث مجموعات على اختبار ما، احسب التباين العام لأداء الافراد ككل.

	العدد (ن)	المتوسط الحسابي	التباين
المجموعة الاولى	30	15	6
المجموعة الثانية	35	20	8
المجموعة الثالثة	25	8	5

الحل: نحسب الوسط الموزون

$$\bar{\text{س}} = \frac{\bar{\text{س}}_1 \times \text{ن}_1 + \bar{\text{س}}_2 \times \text{ن}_2 + \bar{\text{س}}_3 \times \text{ن}_3}{\text{ن}_1 + \text{ن}_2 + \text{ن}_3}$$

$$= \frac{15 \times 30 + 20 \times 35 + 8 \times 25}{30 + 35 + 25}$$

$$= \frac{450+700+200}{90} = \frac{1350}{900} = 15$$

نعوض القيم السابقة بالمعادلة التالية:

$$\text{ع}^2 = \frac{(\text{ن}_1-1)\times\text{ع}_1^2 + (\text{ن}_2-1)\times\text{ع}_2^2 + (\text{ن}_3-1)\times\text{ع}_3^2 + \text{ن}_1(\bar{\text{س}}_1-\bar{\text{س}})^2 + \text{ن}_2(\bar{\text{س}}_2-\bar{\text{س}})^2 + \text{ن}_3(\bar{\text{س}}_3-\bar{\text{س}})^2}{\text{ن}_1+\text{ن}_2+\text{ن}_3-1}$$

$$= \frac{29\times6+34\times8+24\times5+30(15-15)^2+35(20-15)^2+25(10-15)^2}{30+35+25-1}$$

$$= \frac{174+272+120+\text{صفر}+875+625}{90-1} = \frac{2066}{89} \approx 23.21$$

وعليه فإن الانحراف المعياري العام $= \sqrt{\text{ع}^2}$

$$= \sqrt{23.21} = 4.82$$

معامل الاختلاف Coefficient of Variation

يعرف معامل الاختلاف على أنه ناتج قسمة الانحراف المعياري على المتوسط الحسابي مضروباً في100% ، اي النسبة المئوية للانحراف المعياري بالنسبة للمتوسط الحسابي . وبالرموز يعبر عنه بالمعادلة التالية :

$$ف = \frac{ع}{\bar{س}} \times 100\%$$

يستخدم معامل الاختلاف من أجل مقارنة تشتت درجات مجموعتين على متغير ما ؛ اي المقارنة بين اداء مجموعتين لهما نفس وحدات القياس ويختلفان في المتوسط الحسابي . ففي بعض الحالات لا يمكن الاعتماد على المتوسط الحسابي او الانحراف المعياري للحكم على توزيعين معينين من حيث مدى تشتت الدرجات فيهما ، ولا سيما في حالة التوزيعات التكرارية التي تختلف في المتوسطات الحسابية ولها نفس الانحرافات المعيارية ، او تلك التي تتباين في المتوسطات الحسابية والانحرافات المعيارية بشكل ملحوظ . وفي مثل هذه الحالات يتم اللجوء الى معامل الاختلاف لجعل الصورة اوضح من حيث بيان مقدار الشتت بين الدرجات .

يمتاز معامل الاختلاف بأنه لا يعتمد على وحدات القياس المستخدمة للمتغير الاصلي وذلك لانه هو بمثابة حاصل قسمة مؤشرين هي الانحراف المعياري والمتوسط الحسابي ، وبالتالي ليس لهذا المعامل اية وحدات . هذا ويصلح معامل الاختلاف للمقارنة بين تشتت درجات مجموعتين على متغير له نفس وحدات القياس مثل « السم» او « الكم» ولا ينصح باستخدامه في حالة المقارنة بين متغيرات لها وحدات قياس مختلفة .

مثال (5:19) : فيما يلي المتوسطات الحسابية والانحرافات المعيارية لاوزان مجموعتين من الافراد ، قارن بين تشتت الاوزان لكل مجموعة :

المجموعة	المتوسط الحسابي	الانحراف المعياري
المجموعة الاولى	20	6
المجموعة الثانية	18	8

معامل الاختلاف لاوزان المجموعة الاولى = $\frac{6}{20} \times 100\% = 30\%$

معامل الاختلاف لاوزان المجموعة الثانية = $\frac{8}{18} \times 100\% = 44.8\%$

وهكذا تلاحظ ان اوزان المجموعة الثانية اكثر تشتتاً منها في حالة المجموعة الاولى وذلك لان مقدار الاختلاف في المجموعة الثانية =44,4% وهو اكبر من معامل الاختلاف للمجموعة الاولى والبالغة قيمته 30%.

مثال (5:20) : فيما يلي المتوسطات الحسابية والانحرافات المعيارية لاطوال مجموعتين من الافراد.

م1: = المتوسط الحسابي (162) الانحراف المعياري (12)

م2: = المتوسط الحسابي (162) الانحراف المعياري (10)

قارن بين تشتت الاطوال في المجموعتين :

الحل :

معامل الاختلاف للمجموعة الاولى= $\frac{12}{162} \times 100\% = 7.4\%$

معامل الاختلاف للمجموعة الثانية = $\frac{10}{162} \times 100\% = 6.17\%$

نلاحظ ان اطوال المجموعة الثانية اكثر تجانساً منها في المجموعة الاولى وذلك لان معامل الاختلاف للمجموعة الاولى اكبر منه في المجموعة الثانية.

مثال (5:21): فيما يلي المتوسطات الحسابية والانحرافات المعيارية لدرجات مجموعتين من الافراد على اختبار تحصيلي في مادة الرياضيات.

قارن بين تشتت الدرجات لافراد المجموعتين:

م1 : المتوسط الحسابي (22) الانحراف المعياري (8).

م2 : المتوسط الحسابي (24) الانحراف المعياري (8).

الحل:

معامل الاختلاف للمجموعة الاولى= $\frac{8}{22} \times 100\% = 36.4\%$

معامل الاختلاف للمجموعة الثانية = $\frac{8}{24} \times 100\% = 33.3\%$

نجد ان درجات الافراد في المجموعة الاولى أكثر تشتتاً منها في المجموعة الثانية ، فهي اقل تجانساً في المجموعة الاولى مقارنة بدرجات المجموعة الثانية وذلك لان مقدار الاختلاف للمجموعة الاولى اكبر منه للمجموعة الثانية .

الخطأ المعياري:

في بعض الحالات يتم اخذ عينات متعددة كبيرة الحجم من مجتمع احصائي معين بحيث تشتمل كل عينة منها على متوسط حسابي وانحراف معياري للبيانات التي تتضمنها . ومثل هذه المتوسطات يفترض ان تتبع في توزيعها شكل توزيع مماثل لشكل التوزيع لبيانات المجتمع الاحصائي التي سحبت منه العينات . فإذا كان شكل توزيع البيانات في المجتمع الاصلي يأخذ شكل المنحى الطبيعي ، فإن توزيع الاوساط لهذه العينات يتبع شكل التوزيع الطبيعي . اما في حالة كون بيانات المجتمع الاصلي لا تأخذ الشكل الطبيعي ، فعندها يتم اللجوء الى افتراض نظرية النهاية المركزية والذي ينص على أن توزيع اوساط العينات يميل ليأخذ شكل التوزيع الطبيعي بصرف النظر على شكل توزيع البيانات في المجتمع الاصلي .

واعتماداً على ذلك يتم اللجوء في مثل هذه الحالات الى استخدام الخطأ المعياري عوضاً عن الانحراف المعياري وذلك لان هذا الخطأ هو بمثابة تقدير للانحراف المعياري للمتوسطات الحسابية للعينات التي تؤخذ من مجتمع احصائي معين . فالخطأ المعياري هو مؤشر لمدى ابتعاد او اقتراب المتوسطات الحسابية للعينات عن المتوسط الحسابي للمجتمع الاحصائي الاصلي .

وعموماً يتم ايجاد الخطأ المعياري لكل عينة من خلال قسمة الانحراف المعياري للمجتمع على الجذر التربيعي لعدد افراد العينة وذلك كما هو مبين في المعادلة التالية :

$$\text{الخطأ المعياري :} \quad \frac{ع}{\sqrt{ن}}$$

وعليه نلاحظ ان مقدار الخطأ المعياري يعتمد الى درجة كبيرة على حجم العينة من جهة ومقدار الانحراف المعياري لبيانات المجتمع الاحصائي ، بحيث ان هذا الخطأ يزداد بزيادة تشتت بيانات المجتمع الاصلي وانخفاض حجم العينة ، في حين ان قيمته تقل كلما كان حجم العينة كبيراً ومقدار تشتت بيانات المجتمع منخفضاً .

مثال (5:22) : احسب الخطأ المعياري لمتوسط حسابي لعينة حجمها (9) تم سحبها من مجتمع احصائي متوسطه الحسابي= (50) وانحرافه المعياري يساوي (12) .

الحل:

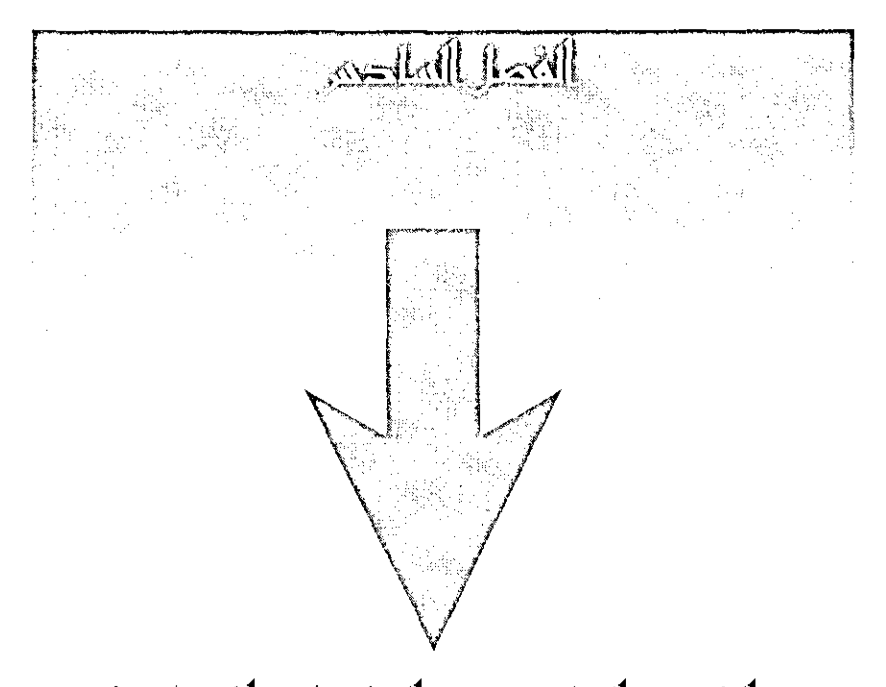
$$\text{الخطأ المعياري} = \frac{12}{\sqrt{9}} = \frac{12}{3} = 4$$

مثال (5:23): احسب الخطأ المعياري لمتوسط حسابي لعينة حجمها (16) تم سحبها من مجتمع احصائي متوسطه الحسابي= (20) وانحرافه المعياري يساوي (4) .

الحل :

$$\text{الخطأ المعياري} = \frac{4}{\sqrt{16}} = \frac{4}{4} = 1$$

لاحظ انه في حالة المثال الاول كان حجم مقدار الخطأ المعياري كبيراً وذلك بسبب ان مقدار التشتت (الانحراف المعياري) لبيانات المجتمع الاصلي كان كبيراً وحجم العينة صغيراً. اما في المثال الثاني فقد كان مقدار الخطأ المعياري صغيراً وذلك لأن مقدار التشتت (الانحراف المعياري) لبيانات المجتمع الاحصائي منخفض وحجم العينة كبير.



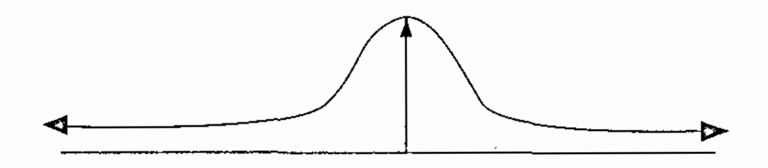
التوزيع الطبيعي والعلامات المعيارية

Normal Distribution & standard Scores

تمـهيـد:

يطلق على هذا التوزيع مسميات أخرى مثل التوزيع الاعتدالي أو السوي أو الجرسي، ويشير إلى تماثل توزيع المشاهدات بين منتصفي متصل غير متناهي الأطراف على نحو تتمركز فيه معظم المشاهدات حول منطقة الوسط لذلك المتصل.

تشيع فكرة استخدام مصطلح التوزيع الطبيعي لدى العاملين في مجالات العلوم الإنسانية والاجتماعيه كافتراض رئيسي لإجراء العمليات الاحصائية والإجابة عن العديد من الأسئلة، وذلك لأن العديد من المتغيرات يقترب شكل توزيعها التكراري من شكل التوزيع المعتدل أو الجرسي، ولا سيما عندما يكون عدد المشاهدات المتعلقة بتلك المتغيرات كبيراً جداً. فلو أخذنا بعض الخصائص والسمات والمتغيرات الإنسانية مثل الذكاء أو الطول أو الوزن أو الدافعية أو التحصيل وغيرها من المتغيرات الأخرى، نجد أن هذه الخصائص أو السمات تتواجد بنسب مختلفة لدى الأفراد، وبالتالي فإن شكل توزيعها التكراري يميل ليأخذ الشكل الاعتدالي أو السوي بحيث يكون الوسط الحسابي مساوياً للوسيط والمنوال، كما هو مبين في الشكل التالي:



شكل (6:1): التوزيع الطبيعي

تكمن أهمية التوزيع الطبيعي في أنه يخلص الباحثين من الكثير من المطالب على نحو يتيح لهم إجراء العديد من العمليات الاحصائية المتقدمة مثل الاحصاء الاستدلالي والتحليلي على البيانات، بالإضافة إلى أنه يعطي معنى لتفسير مفهوم الانحرافات المعيارية. هذا ويمتاز التوزيع الطبيعي بعدد من الخصائص هي:

أولاً: يأخذ الشكل الجرسي: وذلك كما هو مبين في الشكل أعلاه، حيث نلاحظ تمركز معظم المشاهدات في منطقة الوسط، في حين نجد أن عدد المشاهدات يقل تدريجياً كلما اتجهنا إلى مناطق الأطراف. وعليه فإن المشاهدات تتجمع عند المركز (منتصف التوزيع) وتتناقص في مناطق الأطراف.

ثانياً: التماثل: ويتمثل ذلك في تساوي منتصفه الأيمن لمنتصفة الأيسر. فلو نظرنا إلى الشكل أعلاه، نلاحظ أن الخط الرأسي الذي يمر بالوسط الحسابي يقسم التوزيع إلى قسمين متماثلين.

ثالثاً: للتوزيع الطبيعي معلمتان احصائيتان يتحدد في ضونهما شكل التوزيع: وهما المتوسط الحسابي والانحراف المعياري.

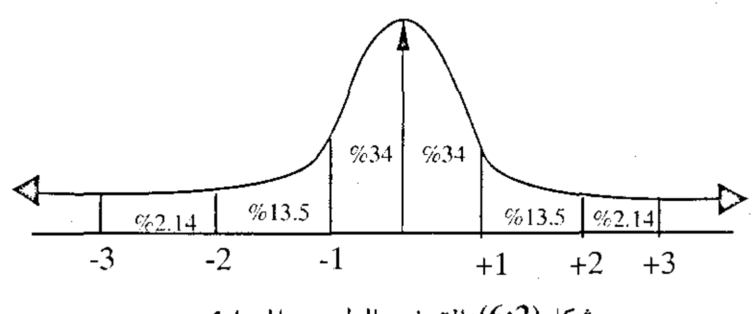
رابعاً: تساوي قيمة المتوسط الحسابي مع قيمة كل من الوسيط والمنوال: حيث أن الوسط = الوسيط = المنوال.

خامساً: التقارب: من حيث أن المشاهدات التي تقع على الأطراف تقترب من المحور الأفقي لكنها لا تتقاطع معه.

سادساً: الاستمرارية والاتصال: ويقصد بذلك أن المشاهدات تنتشر عبر خط متصل غير منتهي الطرفين، بحيث يقل تكرار المشاهدات تدريجياً كلما اتجهنا نحو الطرفين الأيسر والأيمن، الأمر الذي يتيح لنا إمكانية حساب احتمالية (التكرار النسبي) لأي مشاهدة في التوزيع.

سابعاً: نمطية التوزيع: ويقصد بذلك تساوي عدد المشاهدات التي تقع في قسميه الأيمن والأيسر، إذ أن 50% من الحالات تقع في القسم الأيمن ومثلها في القسم الأيسر. واعتماداً على ذلك يمكن حساب النسب المئوية والمئينات وعدد الحالات في أي نقطة من نقاط هذا التوزيع.

وعموماً، نجد أن 68 % من المساحة تقع بين انحراف معياري واحد على يمين المتوسط الحسابي وانحراف معياري واحد على يساره، كما تنحصر 95 % من المساحة في المنطقة الواقعة بين انحرافين معياريين على يمين ويسار المتوسط الحسابي و 99 % من المساحة تقع بين ثلاثة انحرافات معيارية على يمين ويسار المتوسط الحسابي، وذلك كما هو موضح بالشكل التالي:



شكل (6:2): التوزيع الطبيعي المعياري

واعتماداً على معالم هذا التوزيع ، يمكن تحديد المساحات تحت التوزيع الطبيعي والاحتمالات المتوقعة للقيم الاحصائية من خلال حساب المئينات والرتب المئينية لمثل هذه القيم .

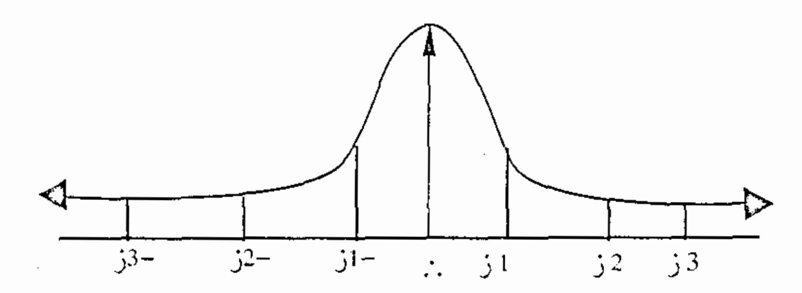
التوزيع الطبيعي المعياري Standard Normal Distribution

يعرف هذا التوزيع أيضاً باسم المنحنى الطبيعي القياسي وهو حالة خاصة من حالات التوزيع الطبيعي بحيث يتم فيه التعامل مع العلامات المعيارية بدلاً من المشاهدات الحقيقية . ففي هذا التوزيع يتم تحويل المشاهدات الخام إلى درجات معيارية تعرف بالعلامات الزائية (Z - scores) من خلال استخدام المعادلة التالية :

$$ز = \frac{س - \bar{س}}{ع}$$

حيث تمثل (س) العلامة الخام ، $\bar{س}$: المتوسط الحسابي ، ع : الانحراف المعياري .

ويمكن القول أن الدرجات المعيارية تتبع التوزيع الطبيعي المعياري إذا كانت المشاهدات الأصلية تتبع التوزيع الطبيعي ، بحيث يكون المتوسط الحسابي لهذا التوزيع صفراً والانحراف المعياري يساوي واحدا ، كما في الشكل التالي :



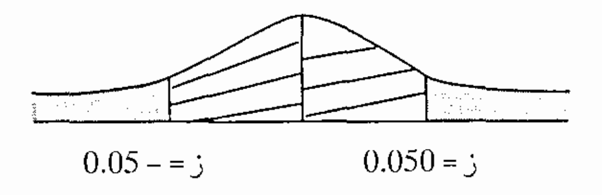
تمثل المساحة الواقعة تحت المنحنى الطبيعي المعياري احتمالات لتوزيع نسب المتغيرات وفق متوسط حسابي معين وانحراف معياري معين ، بحيث تبلغ قيمة المساحة الواقعة تحت المنحنى الطبيعي واحد صحيحا . اعتماداً على هذا يمكن بالتالي من تحديد احتمالية درجة أو علامة معينة في التوزيع . فعلى سبيل المثال . إذا كان أحد المتقدمين لامتحان تحصيلي معين يتمتع بمتوسط حسابي مقداره (50) وانحراف معياري (5) ، فيمكن القول بأن احتمالية أن تتراوح درجة الطالب بين (45 :50 - 5) و (55 : 50 + 5) هو 68% ، وأن احتمالية أن تتراوح درجة الطالب بين (40 :50 - 10) و (60:50+ 10) هو 95 % ، أما احتمالية أن تتراوح درجة الطالب بين (35: 50 - 15) و (65 : 50 + 15) هو 99 % .

لقد قدم الاحصائيون جداول تبين المساحات تحت المنحنى الطبيعي المعياري، بحيث يمكن من خلال هذه الجداول تحديد المساحات التي تقع تحت أو فوق أية درجة معيارية، أو تلك التي تنحصر بين درجتين معياريتين، بحيث يمكن في ضوء هذه المساحة تحديد نسبة أو عدد الحالات المناظرة لتلك المساحة. ومن الجداول المعروفة هناك ما يعرف بجدول التوزيع الطبيعي بدلالة الدرجات المعيارية الزائية. وذلك كما هو مبين في الملحق رقم (2). وفي هذا الجدول توجد الدرجات الزائية وعلى يمينها تقع المساحة المناظرة لها وذلك كما هو موضح بالمثال التالي:

مثال (6:1): نموذج مصغر من الدرجات الزائية والمساحة التي تقع تحتها:

المساحة	الدرجة الزائية (ز)
0.440382	-0.15
0.460172	- 0.10
0.480061	- 0.05
0.50000	0.00
0.519939	0.05
0.539828	0.10
0.559618	0.15

ونظراً لخاصيتي التماثل ونمطية توزيع المساحة تحت المنحى الطبيعي المعياري، نجد أن المساحة التي تقع تحت العلامة الزائية (+ز) هي نفس المساحة التي تقع فوق العلامة الزائية (-ز)، فعلى سبيل المثال، المساحة التي تقع تحت العلامة الزائية (ز = 0.05) هي 0.519939 (انظر الى الجدول اعلاه) وهي نفس المساحة التي تقع فوق العلامة المعيارية (ز = - 0.05) وذلك لأن المساحة التي تقع فوق الدرجة المعيارية (ز = 0.05) هي عبارة عن (1- المساحة التي تقع تحت الدرجة المعيارية (ز =- 0.05))، وبالنظر إلى الجدول نجد أن المساحة تحت (ز =- 0.05) هي 0.480061، وعليه فإن المساحة التي تقع فوقها هي (1- 0.480061 = 0.519939) وهي نفس المساحة التي تقع تحت الدرجة المعيارية (ز = 0.05). والشكل التالي يوضح ذلك

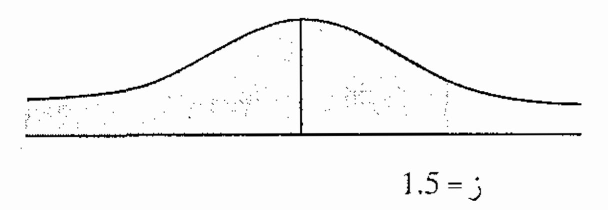


ز = - 0.05 ز = 0.050

في الشكـل أعـلاه لاحظ أن المسـاحـة التي تقع فـوق الدرجـة المعـيارية (+ز) تسـاوي المسـاحـة التي تقـع تحت الدرجـة (-ز)، أي المسـاحـة المضللـة، وعليـه فإن المسـاحـة التي تقـع تحت الدرجـة المعـيـارية (+ز) تساوي المساحة التي تقع فوق الدرجة المعيارية (-ز) وهي المساحة المخططة مضافاً إليها المساحة المضللة.

مثال (6:2) : ما هي المساحة التي تقع تحت الدرجة المعيارية (ز = 1.5) ؟

الحل: المساحة المطلوبة هي المساحة المضللة بالشكل التالي

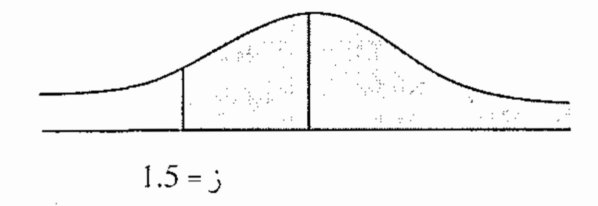


ز = 1.5

بالنظر إلى جدول التوزيع الطبيعي نجد أن قيمة هذه المساحة هي (0.933193)

مثال (6:3) : ما هي المساحة التي تقع فوق الدرجة المعيارية (ز = - 1.5) ؟

الحل: المساحة المطلوبة هي المساحة المضللة بالشكل التالي



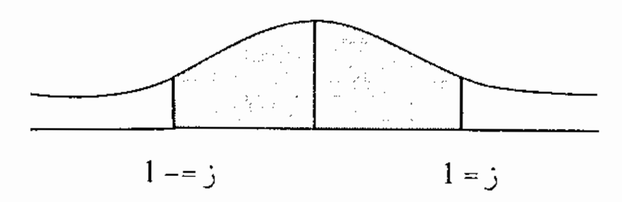
ز = 1.5

وعليه فإن هذه المساحة = المساحة تحت الدرجة المعيارية (ز = +1.5)

ومن الجدول نجدها مباشرة وتساوي (0.933193)

مثال (6:4): ما هي المساحة المحصورة بين الدرجتين المعياريتين (ز = 1) و (ز = -1).

الحل: المساحة المطلوبة هي المساحة المضللة بالشكل التالي :



المساحة المطلوبة = المساحة تحت الدرجة المعيارية (ز = +1) مطروحاً منها المساحة التي تقع تحت الدرجة المعيارية (ز = -1).

المساحة تحت (ز =1) هي: 0.841345

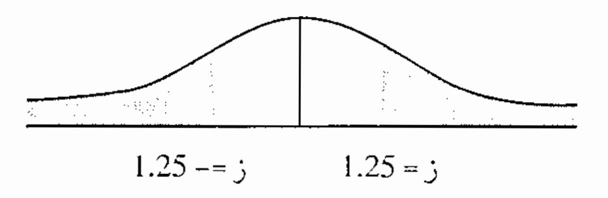
المساحة تحت (ز = -1) هي: 0.158655

وعليه فإن المساحة المطلوبة هي 0.841345 - 0.158655

= 0.682690

مثال(6:5): احسب المساحة التي تقع خارج نطاق الدرجتين المعياريتين (ز = + 1.25) و (ز = -1.25).

الحل: المساحة المطلوبة هي المساحة المضللة بالشكل التالي:



وهذه المساحة هي بمثابة حاصل جمع المساحة التي تقع فوق الدرجة المعيارية (ز = 1.25) وتلك التي تقع تحت الدرجة المعيارية (ز = - 1.25)، أي هي عبارة عن 2 x (المساحة الواقعة تحت الدرجة المعيارية ز = - 1.25).

∴ المساحة تحت الدرجة ز = - 1.25 تساوي (0.10585)، وعليه فإن المساحة المطلوبة هي

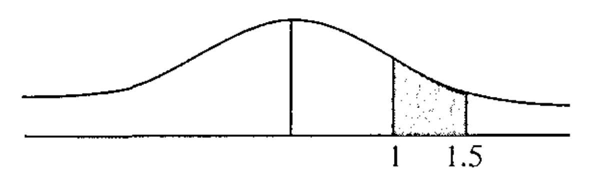
(2 × 0.10585)= 0.211700

وبطريقة أخرى، فإن المساحة التي تقع فوق الدرجة المعيارية ز= 1.25 تساوي المساحة التي تقع تحت الدرجة المعيارية ز = - 1.25، وبما أن المساحة التي تقع فوق الدرجة المعيارية ز = 1.25 تساوي المساحة التي تقع تحت الدرجة المعيارية ز = - 1.25، وعليه فإن المساحة المطلوبة

0.10585 + 0.10585 = 0.211700

مثال (6 : 6): أوجد المساحة المحصورة بين الدرجتين المعياريتين (ز = 1) و (ز = 1.5).

الحل: المساحة المطلوبة هي المساحة المظللة بالشكل التالي:



وهذه المساحة = المساحة الواقعة تحت (ز = 1.5) مطروحاً منها المساحة الواقعة تحت (ز = 1).

المساحة تحت ز = 1.5 = 0.933193

المساحة تحت ز = 1 = 0.841345

المساحة المطلوبة = 0.933193 - 0.841345

= 0.091848

علاقة التوزيع الطبيعي بالمئينات

لقد عرفنا سابقاً أن المساحة التي تقع تحت التوزيع الطبيعي هي عبارة عن احتمال تبلغ قيمته القصوى واحد (1). وعرفنا أيضاً أن هذا التوزيع عبارة عن نقاط أو درجات معيارية يقع دونها نسب أو احتمالات معينة، ولما كان المئين هو بمثابة نقطة في التوزيع يقع دونها نسبة معينة من المشاهدات، وأن الرتبة المئينية هي عبارة عن النسبة المئوية التي تقع دون مشاهدة ما، فعلية فإن المئين هو عبارة عن مقدار العلامة المعيارية (ز) التي يقع تحتها نسبة معينة من الحالات أو مساحة محددة في التوزيع. فعند القول المئين 60، فهذا يشير إلى مقدار العلامة المعيارية (ز) التي يقع تحتها 60 % من المساحة، وعند القول المئين 75، فهذا يشير أيضاً إلى مقدار العلامة المعيارية (ز) التي يقع تحتها 75 % من المساحة.

إما الرتبة المئينية لدرجة معيارية معينة فهي عبارة عن المساحة التي تقع تحت تلك الدرجة، فعند القول أن الرتبة المئينية للدرجة المعيارية ز = 1 هي (0.84)، فهذا يعني ان 84 % من المساحة تقع دون هذه الدرجة المعيارية، وعند القول أن الرتبة المئينية للدرجة المعيارية ز = -1 هي 16 %، فهذا يعني أن 16 % من المساحة تقع تحت العلامة المعيارية ز = -1.

وبما أن التوزيع الطبيعي المعياري ينقسم إلى وحدات متساوية نظراً لخاصية التماثل، فهذا يعني تساوي المسافات التي تفصل المئينات، إذ يمكن القول أن المسافة بين المئين 40 والمئين 50 تساوي المسافة بين المئين 70 والمئين 80، وأن المسافة بين المئين 10 والمئين 40 تساوي المسافة بين المئين 60 والمئين 90.

مثال (6:7): ما هي الدرجة المعيارية (ز) التي تقابل المئين 80؟

الحل: المئين 80 يشير إلى النقطة (الدرجة) التي يقع دونها 80 % من المشاهدات، أي 80 % من المساحة في التوزيع، وبالنظر إلى جدول التوزيع الطبيعي، نجد أن الدرجة المعيارية (ز) التي تقابل هذه المساحة هي ز = 0.85.

مثال (6:8): ما هي الدرجة المعيارية (ز) التي تقابل المئين 40؟

الحل: المئين 40 يشير إلى النقطة (الدرجة) التي يقع دونها 40 % من المساحة. وبالنظر إلى الجدول نجد أن الدرجة المعيارية التي تقابل هذه المساحة هي ز = -0.25.

مثال (6:9): إذا كانت معاملات ذكاء (4000) طالب تتبع شكل التوزيع الطبيعي بمتوسط حسابي مقداره (100) وانحراف معياري قدره (15)، أجب عن الأسئلة التالية:

أ- ما هي نسبة عدد الطلبة الذين معامل ذكائهم أكثر من 115؟

ب- ما هو عدد الطلبة الذين يزيد معامل ذكائهم عن 115؟

ج- ما هو معامل الذكاء الذي يقابل المئين 70

د- كم عدد الطلبة الذين ينحصر معامل ذكائهم بين 85 و 115؟

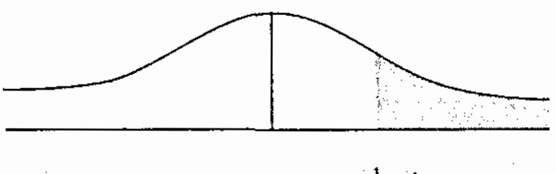
الحل:

أ- نقوم بتحويل العلامة الخام إلى درجة معيارية زائية :

$$ز = \frac{س - \bar{س}}{ع}$$

$$ز = \frac{115 - 100}{15} = \frac{15}{15} = 1$$

وعليه فإن المساحة المطلوبة هي التي تقع فوق هذه الدرجة



ز = 1

ويمكن تحديد هذه المساحة مباشرة من خلال ايجاد المساحة التي تقع دون الدرجة المعيارية (ز =- 1). وبالنظر إلى الجدول نجد أن هذه المساحة = (0.158655). كما يمكن تحديد هذه المساحة على النحو الآتي (1 - المساحة تحت ز = 1). فالمساحة التي تقع تحت ز = 1 هي 0.841345، وعليه فإن المساحة المطلوبة = (1 - 0.841345) = 0.158655 وهي النسبة المطلوبة

ب- عدد الطلبة الذين يزيد معامل ذكائهم عن 115.

كون أن المساحة التي تقع تحت المنحنى الطبيعي هي احتمالات، فعليه يمكن تحديد عدد الحالات التي تقابل هذه الاحتمالات. وبما أن المساحة التي تقع فوق الدرجة المعيارية (ز = 1) تمثل نسبة الحالات التي تقع فوق درجة معامل الذكاء (115)، فعلية فإن عدد الطلاب الذين يزيد معامل ذكائهم عن(115) هو عبارة عن ناتج ضرب المساحة التي تقع فوق الدرجة المعيارية (ز = 1) بعدد الطلاب الكلي.

إذاً عدد الطلبة = 0.158655 x 4000

= 635 طالبا تقريباً

ج- معامل الذكاء الذي يقابل المئين 70.

بما أن المئين هو الدرجة التي تقع دونها نسبة مئوية من عدد الحالات أو مساحة معينه من التوزيع، فعلية فإن تحديد المئين 70 يعني الدرجة المعيارية التي يقع دونها 70 % من المساحة، وبالرجوع إلى الجدول نجد أن الدرجة المعيارية (ز) التي يقع دونها 70 % من المساحة تساوي 0.55 ولتحديد معامل الذكاء المطلوب نستخدم المعادلة:

$$ز = \frac{س - \bar{س}}{ع}$$

وبالتعويض في هذه المعادلة على النحو التالي:

$$0.55 = \frac{س - 100}{15}$$

وبالضرب التبادلي (15 x 0.55) = س - 100

س = 100 + 8.25 = 108 تقريباً.

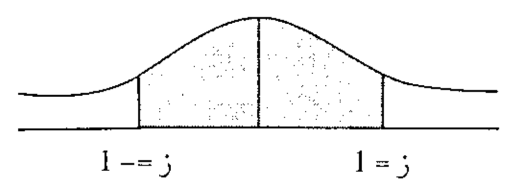
د- كم عدد الطلبة الذين ينحصر معامل ذكائهم بين 85 و 115؟

الحل: يتم تحويل هذه المعاملات إلى درجات معيارية زائية:

$$ز_1 = \frac{100-85}{15} = \frac{-15}{15} = -1$$

$$ز_2 = \frac{100-115}{15} = \frac{15}{15} = +1$$

ثم يتم تحديد المساحة المحصورة بين الدرجتين المعياريتين (ز = - 1) و (ز = 1) :



وهذه المساحة = المساحة تحت (ز = 1) - المساحة تحت (ز = - 1)

= 0.841345 - 0.158655

= 0.68269

وعليه فإن عدد الطلبة المطلوب = 0.68269 × 20000

= 2731 طالبا تقريباً

مثال (6:10): ينتج أحد مصانع الحديد قضبان حديدية من مختلف الأطوال، فإذا كان عدد القضبان التي انتجها في أسبوع واحد (5000) قضيب بمتوسط حسابي مقداره (5 متر) وانحراف معياري قدره (2). أجب عما يلي :

أ- احتمال أن يزيد أطوال القضبان عن 6 متر.

ب- احتمال أن يقل أطوال القضبان عن 3 متر.

ج- عدد القضبان التي يقل طولها عن 4 متر.

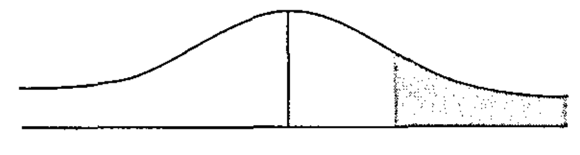
د- عدد القضبان التي يتراوح طولها بين 4 - 7 متر.

الحل:

أ- لتحديد احتمال أن يزيد أطوال القضبان عن 6 متر نحسب الدرجة المعيارية (ز)=

$$ز = \frac{6-5}{2} = \frac{1}{2} = 0.5$$

نحسب المساحة التي تقع فوق الدرجة المعيارية (ز = 0.5)



ز= 0.5

وهذه المساحة = (1 – المساحة تحت الدرجة المعيارية ز = 0.5)

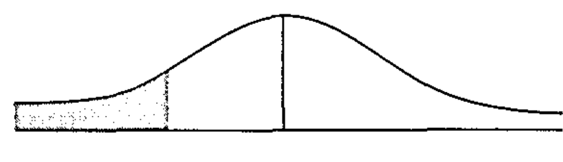
= 1 - 0.691462

= 0.308538 ≈ 31% تقريباً

= أي أن 31 % تقريباً من القضبان يزيد أطوالها عن 6 متر.

ب- لتحديد احتمال أن يقل أطوال القضبان عن 3 متر، نحسب الدرجة المعيارية (ز)

$$ز = \frac{3- 5}{2} = \frac{2}{2} = -1$$



ز = -1

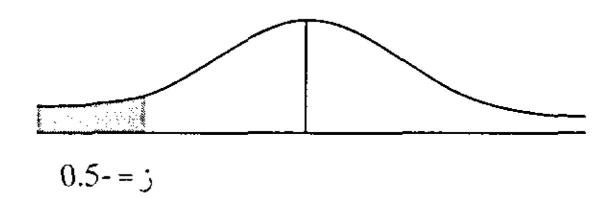
من الجدول نجد أن المساحة الواقعة تحت الدرجة الزائية (-1).

= 0.158655 ≈ 16% تقريباً

أي أن تقريباً 16% من القضبان يقل طولها عن 3 متر.

ج- لحساب عدد القضبان التي يقل طولها عن 4 متر، نحسب الدرجة المعيارية (ز)

$$ز = \frac{4- 5}{2} = \frac{-1}{2} = -0.5$$



ز = -0.5

نجد المساحة الواقعة تحت الدرجة المعيارية (ز = -0.5) وتساوي = 0.308538

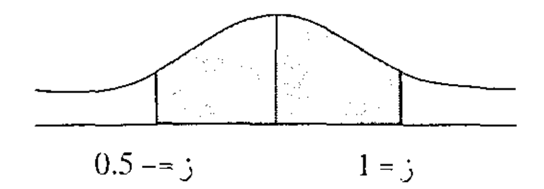
إذا عدد القضبان التي يقل طولها عن 4 متر = 0.308538 × 5000 = 1543 قضيباً تقريباً.

د- لحساب عدد القضبان التي يتراوح طولها بين 4 و 7 متر نحول هذه الدرجات إلى درجات معيارية زائية.

$$ز_1 = \frac{4 - 5}{2} = \frac{-1}{2} = -0.5$$

$$ز_2 = \frac{7 - 5}{2} = \frac{+2}{2} = 1$$

إذاً المساحة المطلوبة هي التي تقع بين الدرجتين المعياريتين (ز = - 0.5) و (ز = 1)



وعليه فإن المساحة المطلوبة هي: المساحة الواقعة تحت الدرجة المعيارية (ز = 1) مطروحاً منها المساحة الواقعة تحت الدرجة المعيارية (ز = 0.5)

المساحة تحت (ز = 1) = 0.84145

المساحة تحت (ز = -0.5) = 0.308538

المساحة المطلوبة 0.841345 - 0.308538

= 0.532807

وعليه فإن عدد القضبان التي يتراوح أطوالها بين 4 متر و 7 متر.

= 0.532807 x 5000 = 2664 قضيبا تقريباً.

مفهوم الدرجات المعيارية:

لقد عرفنا سابقاً أن المنحنى الطبيعي المعياري هو تمثيل بياني للمشاهدات الخام الأصلية بعد تحويلها إلى درجات معيارية تبعاً لمعادلة معينة، وذلك بدلالة المتوسط الحسابي والانحراف المعياري للبيانات الحقيقية. وعليه فالدرجة المعيارية هي بمثابة علامة في توزيع يشكل الانحراف المعياري فيه وحدة قياس يتم في ضوءه المقارنة بين المشاهدات التي يشتمل عليه ذلك التوزيع.

إن تحويل المشاهدات الخام إلى درجات معيارية يوفر لنا معياراً قياسياً موحداً لجميع المشاهدات التي يشتمل عليها توزيع معين. فالمشاهدات الخام تكون عديمة القيمة ولا معنى لها ولا تشكل إلا مجرد رقم ضمن مجموعة معينة، فعلى سبيل المثال، إذا كانت درجة طالب في صف دراسي معين تساوي (75). فمثل هذه الدرجة لا يمكن الحكم عليها ما إذا كانت جيدة أم لا، ما لم يتم تحديد موقعها بالنسبة لدرجات مجموعة الطلاب في الصف. وحتى تصبح لهذه الدرجة معنى يجب مقارنتها مع درجات مجموعة الطلاب في الصف الذي ينتمي إليه.

من جهة أخرى، لا يمكن بأي شكل من الأشكال المقارنة بين درجتين في توزيعين مختلفين اعتماداً على البيانات الخام، فعلى سبيل المثال إذا كانت درجة طالب في امتحان العلوم = 70، ودرجته في امتحان الرياضيات = 65، فلا يمكن القول أن تحصيله في مادة العلوم أفضل من تحصيله في مادة الرياضيات من مجرد النظر إلى الدرجات الخام، وذلك لأن كل درجة من هذه الدرجات تنتمي إلى مجموعة معينة مختلفة المعالم الاحصائية.

وعليه، فحتى يتسنى لنا إمكانية الحكم على المشاهدات في توزيع معين وتفسيرها على نحو يبدو ذو دلالة ومعنى، وحتى نستطيع مقارنة كل مشاهدة في التوزيع مع غيرها من المشاهدات في ذلك التوزيع أو مقارنة مشاهدات في توزيعات احتمالية مختلفة، يجب تحويل المشاهدات الخام إلى صيغ جديدة وفق أحد نوعين من التحويل هما: التحويل الخطي والتحويل غير الخطي.

ففي التحويل الخطي يتم تحويل المشاهدات الخام إلى درجات معيارية وفق معادلات خاصة بحيث تكون العلاقة بين المشاهدات الخام الأصلية والدرجات المعيارية المقابلة لها خطية. وفي مثل هذا النوع من التحويل يتم الاحتفاظ بشكل التوزيع للمشاهدات الخام الأصلية، أي بمعنى أن توزيع الدرجات المعيارية يتخذ نفس شكل توزيع المشاهدات الخام. يتضمن التحويل الخطي نوعين من الدرجات المعيارية هما الدرجة المعيارية الزائيه والدرجة المعيارية التائيه، ومثل هاتين الدرجتين توجد علاقة رياضية تجمع بينهما.

أولاً: الدرجة المعيارية الزائية (ز:Z):

تعرف الدرجة المعيارية الزائية على أنها علامة في توزيع معياري وسطه الحسابي = صفر وانحرافه المعياري = 1. وهي تمثل مقدار انحراف المشاهدة الخام الأصلية على وسطها الحسابي بالنسبة إلى الانحراف المعياري، وذلك وفق المعادلة التالية: $ز = \frac{س - \bar{س}}{ع}$

توفر لنا الدرجة المعيارية الزائيه معياراً قياسياً موحداً لتحديد وضع المشاهدات الخام الأصلية بالنسبة للمجموعة التي تنتمي إليها، كما أنها تتيح لنا إمكانية المقارنة بين تلك المشاهدات والحكم عليها. فالدرجة المعيارية الموجبة تمثل عدد الانحرافات المعيارية التي تزيد بها الدرجة الخام الأصلية عن الوسط الحسابي للمجموعة، في حين نجد ان الدرجة المعيارية السالبة تمثل عدد الانحرافات المعيارية التي تقل بها الدرجة الخام الأصلية عن الوسط الحسابي للمجموعة. وتحديداً يمكن تفسير الدرجة المعيارية الزائية على النحو التالي:

1- عند القول أن الدرجة المعيارية (ز = +1)، فهذا يعني أن المشاهدات الخام تزيد عن المتوسط الحسابي بمقدار انحراف معياري 1. وهذا يعني أن هذه المشاهدة أفضل من 84 % من مجموع المشاهدات.

2- إذا كانت الدرجة المعيارية (ز = - 1)، فهذا يشير إلى أن المشاهدة الأصلية الخام تقل عن المتوسط الحسابي للمجموعة بمقدار انحراف معياري واحد. وهذا يشير إلى أن هذه المشاهدة أفضل من 16 % من مجموع المشاهدات الكلي.

3- إذا كانت الدرجة المعيارية (ز = صفر)، فهذا يعني أن المشاهدة الخام الأصلية تساوي المتوسط الحسابي لمجموع المشاهدات. ومثل هذه المشاهدة تقع في مركز التوزيع بحيث تعد أفضل من 50 % من مجموع المشاهدات.

ولجعل الأمر أكثر وضوحاً، يمكن الاستعانة بالأمثلة التالية:

مثال (6:11): لقد حصل طالب على درجة (55) في امتحان لمادة العلوم وكان المتوسط الحسابي لدرجات أفراد المجموعة يساوي (60) والانحراف المعياري =(2). فما هو وضع الطالب بالنسبة للمجموعة:

الحل:

$$= \frac{60-55}{2} = \frac{5-}{2} = 2.5-$$

المساحة التي تقع تحت هذه الدرجة تساوي 0.00621

يلاحظ أن تحصيل الطالب في هذه المادة ضعيف جداً إذا ما قورن مع تحصيل بقية الطلاب في المجموعة لهذه المادة، حيث أن درجته هي أفضل من 0.06% من مجمل الدرجات.

مثال (6:12): فيما يلي بعض البيانات المتعلقة بتحصيل طالب في أربعة مواد على النحو الآتي:

قارن بين درجات الطالب على هذه المواد مرتباً مستوى الأداء لهذا الطالب على المواد تنازلياً؟

المادة	الدرجة	المتوسط الحسابي لدرجات المجموعة	الانحراف المعياري
الرياضيات	72	70	2
الفيزياء	80	90	5
الكيمياء	65	55	4
الأحياء	84	75	3

الحل : حتى نتمكن من إجراء المقارنة يجب تحويل الدرجات الخام إلى درجات معيارية زائيه وذلك من أجل ايجاد معيار قياسي موحد للمقارنة .

$$1 = \frac{70\text{-}72}{2} = ز_1$$

المساحة المقابلة للدرجة المعيارية (ز = 1) = 0.841345

وهذا يعني أن تحصل الطالب في مادة الرياضيات أفضل من تحصيل 84 % من مجموع الطلاب في المجموعة .

$$2\text{-} = \frac{10\text{-}}{2} = \frac{90\text{-}80}{5} = ز_2$$

المساحة المقابلة لها = 0.00621

وهذا يعني أن تحصيل الطالب في مادة الفيزياء منخفض جداً وهو أفضل فقط من تحصيل 0.06% من مجموع الطلاب .

$$2.5 = \frac{10}{4} = \frac{55\ \text{-}65}{4} = ز_3$$

المساحة المقابلة لها = 0.99379

وهذا يعني أن تحصل الطالب في مادة الكيمياء أفضل من تحصيل 99.4% من مجمل الطلاب .

$$3 = \frac{9}{3} = \frac{75\text{-}\ 84}{3} = ز_4$$

المساحة المقابلة لها = 0.99865

وهذا يعني أنه تحصل الطالب في مادة الأحياء أفضل من تحصيل 99.9 % من مجمل الطلاب .

وعليه يمكن ترتيب تحصيل الطالب تنازلياً على هذه المواد من المستوى الأعلى إلى الأقل وذلك على النحو الآتي الأحياء، الكيمياء، الرياضيات، ثم الفيزياء، وبذلك نجد أن أفضل تحصيل لهذا الطالب كان في مادة الأحياء ثم الكيمياء فالرياضيات وأخيراً الفيزياء.

ثانياً: الدرجة المعيارية التائيه (ت: t)

تعد الدرجة التائية أحد أشكال التحويل الخطي للبيانات الخام وفق متوسط حسابي مقداره (50) وانحراف معياري قدره (10). ويفضل اللجوء إلى الدرجة التائية للتخلص من القيم السالبة أو الكسور العشرية التي قد تتخذها الدرجة المعيارية الزائيه. وهناك مبرر آخر يبرر استخدام الدرجة التائيه هو أن غير المختصين في الاحصاء عادة لا يدركون ماذا تعني العلامات المعيارية، فعند القول على سبيل المثال لولي أمر أحد الطلاب أن درجة ابنه المعيارية في التحصيل تساوي صفراً، فهذا ربما يسبب صدمة لولي أمر هذا الطالب، حيث أنه لا يدرك أن هذه الدرجة تساوي المتوسط الحسابي لدرجات مجمل الطلاب.

يتطلب ايجاد القيمة التائيه تحويل الدرجات الخام إلى قيم زائيه ثم استخدام المعادلة التالية:

ت = 10 × ز + 50.

هذا ويعتمد شكل توزيع الدرجات التائيه على شكل التوزيع للدرجات الخام الأصلية، فإذا كان شكل التوزيع للدرجات الخام يتبع التوزيع الطبيعي، فإن شكل التوزيع للدرجات التائية يتخذ شكل المنحنى الطبيعي. ومن الجدير ذكره هنا، أن قيم التوزيع التائي تتراوح بين صفر و (100) حيث أن ما يقارب من 99 % من مجمل الدرجات تقع ضمن ثلاثة انحرافات معيارية.

مثال (6:13): إذا كانت الدرجة الزائيه لعلامة طالب (ز = 2)، فما هي الدرجة التائية لعلامة الطالب؟

الحل: ت = 10 × ز + 50

= 10 × 2 + 50 = 70

مثال (6:14): ما هي الدرجة التائية التي تقابل الدرجة الزائية (ز = -0.5)؟

الحل: ت= 10 × (-0.5) + 50

= -5 + 50 = 45

مثال (6:15): ما هي العلامة الزائية التي تقابل الدرجة التائية 85 ؟

الحل: ت = 10 × ز + 50

85 = 10 ز + 50

85 -50 = 10 ز .

35 = 10ز

ز = $\frac{35}{10}$ = 3.5

مثال (6:16) : ما هي العلامة الزائية التي تقابل الدرجة التائية 20 ؟ .

ت = 10 × ز + 50

20 = 10 ز + 50

10 ز = 20 - 50

10 ز = - 30

∴ ز = -3

مثال(6:17) : ما العلامة التائية التي تقابل الدرجة الزائية (ز = 1.5) .

ت = 10 × ز + 50

= 10 x 1.5 + 50

= 15 + 50

= 65

هناك طريقة أخرى لحساب الدرجة المعيارية ولا سيما في حالة التوزيعات التكرارية ذات الفئات المفتوحة أو في حالة بعض التوزيعات التي يصعب فيها ايجاد المتوسط والانحراف المعياري، بحيث يتم الاستعاضة عنهما واستخدام كل من الوسيط والمدى الربعي بدلاً منهما لحساب الدرجة المعيارية وذلك وفق المعادلة التالية :

$$\text{الدرجة المعيارية} = \frac{\text{الدرجة الخام} - \text{الوسيط}}{\text{المدى الربعي}}$$

مثال(6:18) : في توزيع تكراري ذي فئات مفتوحة كان الوسيط لهذا التوزيع يساوي (70) وكان المدى الربعي له = 20، احسب الدرجة المعيارية للدرجة (80) .

الحل: الدرجة المعيارية = $\frac{80 - 70}{20}$ = $\frac{50 - 10}{20}$ = 0.5

مثال(6:19) : إذا كان الوسيط لتوزيع معين = 50 والمدى الربعي لهذا التوزيع = 15 ، فما هي الدرجة المعيارية للعلامة (50)

الحل: ز = $\frac{50-50}{15}$ = صفر

مثال(6:20) : إذا كان الوسيط لتوزيع معين = 50 والمدى الربعي لهذا التوزيع = 15 ، فما هي الدرجة المعيارية للعلامة (40)

الحل: = الدرجة المعيارية = $\frac{50-40}{15}$ = $\frac{-10}{15}$ = -0.67

التحويل غير الخطي للبيانات

عرفنا سابقاً أن التحويل الخطي للبيانات ينطوي على إضافة مقدار ثابت إلى كل قيمة من بيانات التوزيع أو طرح مقدار ثابت منها أو ضربها أو قسمتها على مقدار ثابت ، ومثل هذا التحويل لا يؤثر على شكل التوزيع للبيانات الأصلية ، إذ أن البيانات الجديدة تتبع نفس شكل التوزيع للبيانات الأصلية .

ولكن نجد أن التحويل غير الخطي يؤدي إلى التغيير في شكل البيانات الجديدة وذلك لأن البيانات الأصلية يتم تحويلها بدلالة جذرها أو لوغريتمها . وفي مثل هذه الحالات عادة يتم اللجوء إلى استخدام الدرجة المعيارية الزائية المعدلة أو العلامة المئينية التساعية .

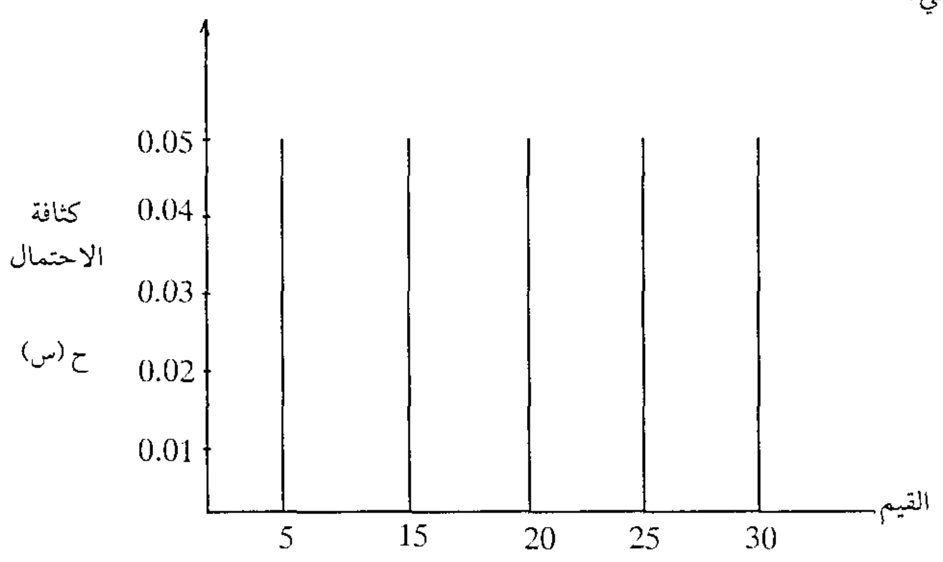
وتجدر الإشارة هنا أن شكل التوزيع للبيانات بعد التحويل يتم تعديله ليأخذ المنحنى الطبيعي على افتراض أن السمات الإنسانية عموماً تتجه لتأخذ شكل التوزيع الطبيعي .

فالدرجة الزائيه المعدلة هي بمثابة علامة معيارية في توزيع وسطه الحسابي يساوي صفراً وقيمة انحرافه المعياري تساوي (1) ، ومثل هذه الدرجة يتم ايجادها من خلال تحويل العلامات الخام إلى رتب مئينية ومن ثم ايجاد العلامة الزائية المقابلة لها من جدول التوزيع الطبيعي .

أما العلامة التساعية فهي أيضاً علامة معيارية في توزيع يشتمل على تسع فئات تسمى تساعيات بحيث يشتمل كل تساعي (فئة) منها على نسبة معينة من الحالات (البيانات) . ويتم ايجاد الدرجة التساعية من خلال تحويل كل علامة خام في التوزيع إلى رتبة مئينية وتحديد الفئة التي تنتمي إليها في الجدول التساعي .

التوزيع المستطيل:

يعد التوزيع المستطيل أحد التوزيعات غير الطبيعية للبيانات بحيث يأخذ توزيع البيانات شكلاً مستطيلاً وفيه تتساوى التكرارات لجميع القيم في ذلك التوزيع ، أي أن توزيع المتغير موضع القياس يكون منتظماً من حيث استواء عدد التكرارات لكل قيمة من قيم هذا المتغير، وذلك كما في الشكل التالي :



يمتاز التوزيع المستطيل بكونه متماثل حول الوسط الحسابي ويمتلك معلمتين رئيستين هما الحد الأدنى (ع) والحد الأعلى (ص) للمتغير، وفيه يتساوى الوسط مع الوسيط في الوقت الذي لا توجد له قيمة منوال. وفي حالة التوزيع المستطيل تسهل عملية تحديد المساحات تحت المنحنى لأن المساحات هي عبارة عن مستطيلات يمكن ايجاد مساحتها من حاصل ضرب القاعدة في الارتفاع.

تستخدم عادة معادلة خاصة لتحديد ما يسمى بكثافة الاحتمال لقيم المتغير في التوزيع المستطيل وهي :

$$ح(س) = \frac{1}{ص - ع}$$

حيث (ع : الحد الآدنى للمتغير)،

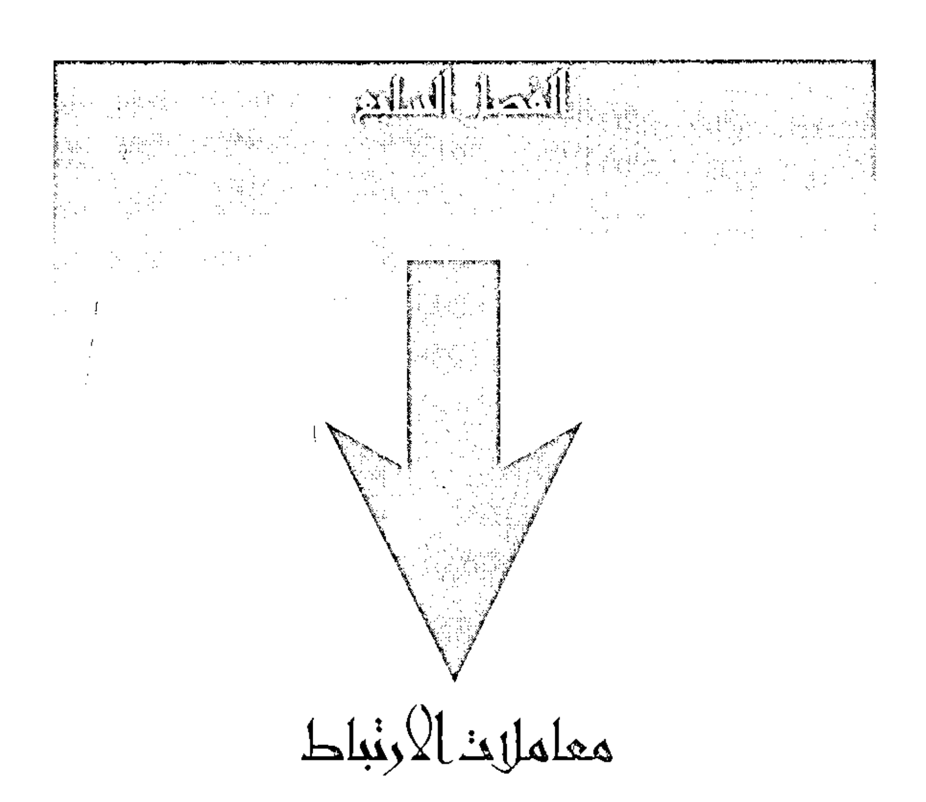
(ص : الحد الآعلى للمتغير)،

أما الوسط لهذا التوزيع فيتم ايجاده من خلال المعادلة التالية :

$$\bar{س} = \frac{ص + ع}{2}$$

في حين يتم حساب التباين من خلال المعادلة التالية :

$$التباين = \frac{(ص - ع)^2}{2}$$



معاملات الارتباط

Correlation Coefficients

تمهيد:

خلال الفصول السابقة تم التعرض للمعالجات الاحصائية للبيانات التي تتعلق بمتغير احصائي واحد، من حيث بيان شكل التوزيع لقيم هذا المتغير، أو تحديد بعض القيم الاحصائية ذات العلاقة مثل المنوال والوسيط والمئينات والمتوسط الحسابي والانحراف المعياري والدرجات المعيارية، ولكن في بعض الحالات يتطلب الأمر جمع بيانات حول مجتمع احصائي تتعلق باكثر من متغير واحد، أو بيانات مختلفه يتم جمعها حول متغير احصائي واحد في اوقات أو ظروف مختلفة، ويسعى الباحث إلى الكشف عن درجة الترابط بين تلك البيانات أو تحديد العلاقة بين متغيرين أو أكثر بالنسبة لذلك المجتمع الاحصائي. ففي مثل هذه الحالات، نجد ان المعالجات الاحصائية التي ورد ذكرها سابقا لا تفي بالغرض المطلوب لان مثل هذه المعالجات تصلح للتعامل مع بيانات ترتبط بمتغير احصائي واحد. وعليه فلابد من البحث عن احصائي مناسب يمكن من الكشف عن العلاقات القائمة بين المتغيرات، و بذلك نجد ان معامل الارتباط يعد الاحصائي المناسب الذي يخدم ذالك الغرض. يعد معامل الارتباط مؤشرا احصائيا يكشف عن وجود أو عدم وجود علاقة بين متغيرين أو ظاهرتين، وهو يعبر عن قوة واتجاه العلاقة بين هذين المتغيرين. بالاضافة لذلك قد يعبر معامل الارتباط عن قوة واتجاه العلاقة بين اكثر من متغيرين وذلك كما هو الحال في دراسة العلاقات بين عدد من المتغييرات الاحصائية بالوقت نفسه.

تجدر الاشاره هنا، ان وجود علاقة بين متغيرين لا يعني بالضرورة وجود علاقة سببيه بينهما، أي لا يعني ان احد المتغيرين هو سبب في حدوث الاخر. ولكن يتضمن ذلك ان وجود احدهما ربما يؤدي إلى زيادة أو نقصان الاخر. فعلى سبيل المثال نلاحظ ان الاشجار تنمو بصرف النظر ان وجود السماد أو عدمه، ولكن في حال توفير السماد لها نلاحظ ان معدل نموها يزداد، وهذا يفسر وجود علاقة ارتباطيه موجبة بين توفير السماد ومستوى النمو في البنباتات.

يمكن لمعامل الارتباط ان يأخذ قيمة ضمن مدى يتراوح بين (-1) و (+1) مرورا بالصفر، وقد يأخذ قيمة سالبه أو موجبه. ففي حالة كون معامل الارتباط سالباً، فهذا يشير إلى وجود علاقة عكسيه بين

المتغيريين ، إذ ان زيادة احد المتغيرين يؤدي إلى نقصان في المتغير الاخر؛ أي أن المتغيرين يسيران في اتجاهين متضادين . اما في حالة كونه موجباً ، فعندها يعد معامل الارتباط مؤشرا لوجود علاقة طردية بين المتغيرين ، حيث ان الزيادة في احد المتغيرين تتبعه بالضرورة زيادة في المتغير الاخر ، وهذا يعني انهما يسيران في نفس الاتجاه .

وانطلاقا من ذلك فأن الاشارة الموجبة أو السالبة لمعامل الارتباط لا تدل بأي شكل من الاشكال على قوة العلاقة بين المتغيرات ، ولكنها هي مؤشر لاتجاه العلاقة . اما قوة العلاقة فتقاس من خلال القيمة المطلقة لمعامل الارتباط . ففي حال كون قيمة معامل الارتباط بين متغيرين تساوي (± 1) أو تقترب من ذلك ، فعندها يقال بأن العلاقة بين المتغيرين علاقة تامه وخطيه أو شبه خطيه ، ولكن في حال كون قيمة معامل الارتباط تساوي (صفرا) أو تقترب من ذلك ، فعندها يمكن القول بعدم وجود علاقة بين المتغيرين ، إذ أن التغير في احد المتغيرين لا يتأثر أو يؤثر بالتغيرات التي تطرأ على المتغير الاخر ، ومن الامثلة على هذا النوع من العلاقات ، العلاقة بين اللون والتحصيل الدراسي أو لون العيون ودرجة الذكاء حيث لا يوجد أية علاقة بين هذه المتغيرات . هذا ويلخص الجدول التالي الدلالات المرتبطة بمعامل الارتباط من حيث قوة واتجاه العلاقة .

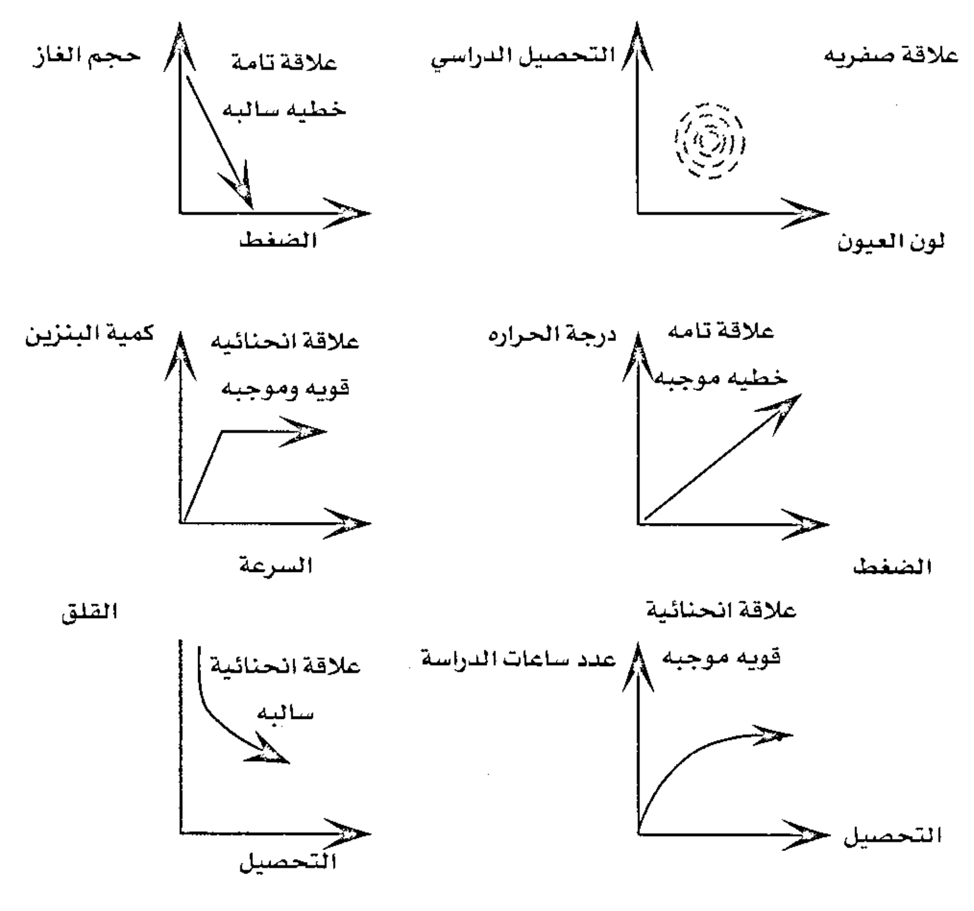
جدول رقم (7:1): الدلالات المرتبطه بقوة واتجاه الارتباط)

مدى قيم معامل الارتباط	قوة العلاقة	اتجاه العلاقة
(+1) أو (−1)	علاقة تامه	طردية (+1) عكسيه (−1)
± 0.99 − ± 0.60	قويه	(+) طرديه (−) عكسيه
± 0.59− ± 0.40	متوسطه القوة	(+) طرديه (−) عكسيه
± 0.39 − ± 0.01	ضعيفه	(+) طرديه (−) عكسيه
صفر	عدم وجود علاقة	علاقة صفريه

يندر ايجاد علاقة تامه بين المتغيرات في العلوم الانسانية وذلك لان معظم المتغيرات ترتبط بخصائص انسانية ومثل هذه المتغيرات مترابطة ومتداخلة معا ، إذ يصعب عملية ضبطها أو عزلها عن بعضها . وعموما يمكن الحصول على معاملات ارتباط مختلفة القوة والاتجاه بين المتغيرات المتعددة في العلوم الإنسانية. ومن الامثلة على العلاقات الموجبه القوية، العلاقة بين عدد ساعات الدراسة والتحصيل الدراسي، أو العلاقة بين معامل الذكاء والتحصيل الدراسي. اما في حالة العلاقات السالبه القوية فهناك العلاقة بين مستوى القلق والتحصيل الدراسي، إذ أن مستوى التحصل يقل مع زيادة مستوى القلق لدى المتعلم. هذا ويمكن الاستفادة من دراسة معاملات الارتباط لخدمة عدد من الاغراض تتمثل في الاتي :

1- الكشف عن قوة العلاقة بين المتغيرات من حيث كونها قوية أو ضعيفه أو متوسطه أو صفرية .

2- تحديد شكل العلاقة بين المتغيرات من حيث كونها خطيه أو انحنائيه أو غير منتظمه. وبالنظر إلى الاشكال التالية : نلاحظ الانماط المختلفة من العلاقات



3- تحديد اتجاه العلاقة بين المتغيرات من حيث كونها طردية أو عكسية أو محايدة (صفرية).
4- التنبوء بقيمة احد المتغيرات من خلال تحديد قيم متغير اخر يرتبط به.

وهذا يعني بالطبع امكانية التنبوء بحدوث الظواهر المختلفة من خلال بروز المؤشرات الدالة عليها أو تلك التي ترتبط معها بعلاقة معينه.

- **انواع الارتباطات Types of Correlations**

يمكن تقسيم الارتباطات إلى عدد من الانواع اعتمادا على عدد ونوعيه المتغيرات الداخلة فيها، وهذه الانواع هي:

1- الارتباط البسيط (Simple Correlation): وهو الارتباط الذي يبحث العلاقة بين متغيرين أو ظاهرتين فقط.
2- الارتباط المتعدد (Multiple Correlation): وهو الارتباط الذي يبحث العلاقة بين متغير واحد وعدد من المتغيرات يتراوح عددها بين متغيرين أو أكثر.
3- الارتباط الجزئي (Partial Correlation): وهو الارتباط الذي يبحث العلاقة بين مجموع المتغير الكلي وعنصر أو جزء من ذلك المتغير مثل العلاقة بين الدرجة الكيلة على اختبار تحصيلي معين وعلامة سؤال واحد من ذلك الاختيار، كما ان هذا النوع الارتباط يبحث ايضا العلاقة بين متغيرين فقط من مجموعة متغيرات على اعتبار ان بقية المتغيرات ثابتة لا اثر لها.

- **طرق حساب معامل الارتباط:**

هناك عدة طرق لحساب معاملات الارتباط تبعا لعدد وطبيعة المتغيرات المراد دراسة العلاقة بينها. وفيما يلي عرض للطرق المختلفة المستخدمة في حساب معاملات الارتباط:

أولا: معامل ارتباط بيرسون Pearson Correlation

تعد طريقة حساب معامل الارتباط من خلال معامل بيرسون من أكثر طرق شيوعا واستخداما ولا سيما في حالة المتيغرات المتصله؛ فهو يستخدم لحساب العلاقات بين متغيرين يقعان على مستوى القياس الفئوى أو النسبي، أو بين متغيرين احدهما يقع على المستوى النسبي والأخر على المستوى الفئوي.

عند استخدام هذه الطريقة هناك عدة افتراضات يتطلب توفرها تتمثل في توفرعلاقة خطيه بين المتغيرات، وتجانس التباين، اي ثبات تباين قيم احد المتغيرين عند كل قيمة من قيم المتغير الاخر والعكس بالعكس. وعموماً عند استخدام هذه الطريقة غالبا يفترض الباحث مسبقا تحقق مثل هذه الافتراضات في حالة المتغيرين المنوي تحديد العلاقة بينهما.

هناك عدة طرق يمكن استخدامها لحساب معامل ارتباط بيرسون تتمثل في ما يأتي :

أ) طريقة الانحرافات .

ب) طريقة البيانات الاصليه .

ج) طريقة التغاير .

د) طريقة الدرجات المعيارية .

عموما تعد طريقة الانحرافات وطريقة البيانات الاصلية هي الاكثر شيوعاً واستخداما في حساب معامل ارتباط بيرسون لانها تتعامل مع المشاهدات الاصلية ولا تتطلب اجراء اية انواع من التحويل أو التغيير على مثل هذه المشاهدات ، وفي هذا الفصل سوف يتم التعرض لهاتين الطريقتين فقط .

أولا: طريقة الانحرافات:

تتطلب هذه الطريقة تحديد انحرافات القيم عن متوسطها الحسابي ومن ثم ايجاد مربعات هذه الانحرافات وتعويض القيم بالمعادلة التالية :

$$ر_{س ص} = \frac{\sum (س - \bar{س})(ص - \bar{ص})}{\sqrt{\sum (س - \bar{س})^2 (ص - \bar{ص})^2}}$$

مثال (7:1): في الجدول ادناه بيانات حول متغيرات (س ، ص) احسب معامل ارتباط بيرسون بين هذين المتغيرين .

الرقم	س	ص	س - سَ	ص - صَ	(س - سَ) (ص - صَ)	$(س - \bar{س})^2$	$(ص - \bar{ص})^2$
1	2	1	-4	-6	24	16	36
2	3	4	-3	-3	9	9	9
3	5	7	-1	صفر	صفر	1	صفر
4	7	8	+1	+1	1	1	1
5	9	10	+3	+3	9	9	9
6	10	12	+4	+5	20	16	25
المجموع	36	42	صفر	صفر	63	52	80

الحل: 1- نحسب اولا المتوسط الحسابي لقيم (س) والمتوسط الحسابي لقيم (ص) .

$$\bar{س} = \frac{\sum س}{ن} = \frac{36}{6} = 6$$

$$\bar{ص} = \frac{\sum ص}{ن} = \frac{42}{6} = 7$$

2- نحسب انحرافات القيم عن متوسطها الحسابي لقيم (س) وقيم (ص)، وذلك كما يظهر في العمودين الثالث والرابع في الجدول اعلاه.

3- نجد ناتج ضرب انحرافات قيم (س) عن متوسطها الحسابي بانحرافات قيم (ص) عند متوسطها الحسابي كما هو موضح في العمود الخامس.

4- نجد مربعات انحرافات قيم (س) عن متوسطها الحسابي وكذلك مربعات انحرافات قيم (ص) عن متوسطها وذلك كما هو الحال في العمودين السادس والسابع في الجدول اعلاه.

5- نعوض القيم في المعادلة التالية:

$$ر_{س ص} = \frac{\sum (س - \bar{س})(ص - \bar{ص})}{\sqrt{\sum (س - \bar{س})^2 (ص - \bar{ص})^2}} = \frac{63}{\sqrt{80 \times 52}} = \frac{63}{\sqrt{4160}}$$

$$= \frac{63}{64.5} = 0.97 \text{ تقريباً}$$

ثانيا: طريقة البيانات الاصلي:

يمكن اللجوء إلى هذه الطريقة لتجنب الصعوبات الناجمة عن وجود كسور عشرية ولا سيما في حالة كون المتوسط الحسابي يشتمل على كسر عشري، بحيث يتم التعامل مباشرة مع البيانات الاصلية دون الحاجة إلى ايجاد انحرافاتها عن المتوسط الحسابي وذلك وفق المعادلة التالية:

$$ر_{س ص} = \frac{ن \sum س ص - \sum س \times \sum ص}{\sqrt{\left[ن \sum س^2 - (\sum س)^2\right]\left[ن \sum ص^2 - (\sum ص)^2\right]}}$$

مثال (7:2) : أحسب معامل ارتباط بيرسون بين علامات مجموعة من الطلاب على مادة الرياضيات (س) ومادة الفيزياء (ص) في الجدول ادناه :

الرقم	س	ص	$س^2$	$ص^2$	س × ص
1	4	2	16	4	8
2	7	3	49	9	21
3	8	6	64	36	48
4	10	7	100	49	70
5	15	10	225	100	150
المجموع	44	28	454	198	297

الحل:

1- نحسب مجموع قيم (س) وقيم (ص) كما هو مبين في اسفل العمودين الثاني والثالث

$\sum س = 44$ ، $\sum ص = 28$

2- نجد مربع قيم (س) ومربع قيم (ص) كما هو مبين في العمودين الرابع والخامس.

$\sum س^2 = 454$ ، $\sum ص^2 = 198$

3-نحسب ناتج ضرب قيم (س) بقيم (ص) كما هو مبين في العمود السادس

$\sum س \times ص = 297$

4- نحسب ناتج ضرب مجموع قيم (س) بمجموع قيم (ص)

$\sum س \times \sum ص = 44 \times 28 = 1232$

5- نعوض القيم بالمعادلة التالية :

$$ر_{س ص} = \frac{ن \sum س ص - \sum س \times \sum ص}{\sqrt{\left[ن \sum س^2 - (\sum س)^2\right]\left[ن \sum ص^2 - (\sum ص)^2\right]}}$$

$$ر_{س ص} = \frac{5(297) - 1232}{\sqrt{[5(454) - (454)^2][5(198) - (28)^2]}}$$

$$ر_{س ص} = \frac{1485 - 1232}{\sqrt{[2270 - 1936][990 - 784]}}$$

$$ر_{س ص} = \frac{253}{\sqrt{(334)(206)}}$$

$$= \frac{253}{\sqrt{68804}}$$

$$= \frac{253}{262.30} = 0.96 \text{ تقريباً}$$

ثانيا: معامل ارتباط سبيرمان Spearman's Correlation Coefficient:

يطلق على معامل الارتباط هذا اسم معامل ارتباط الرتب بحيث يستخدم في حالة كون المتغيرات المراد حساب معامل الارتباط بينها تقع ضمن مستوى مقياس الرتب؛ أي في حالة المتغيرات الرتبيه. ففي مثل حالة هذه المتغيرات يتم حساب معامل الارتباط بين رتب القيم وليس بين القيم الاصلة لتلك المتغيرات. عموما يمتاز معامل الارتباط هذا بسهوله حسابه بالرغم انه قيمته لا تساوي قيمة معامل الارتباط المحسوب بين القيم الاصلية للمتغيرات، إذ ان معامل الارتباط المحسوب بين القيم الاصلية هو في الغالب اكثر دقة من ذلك المحسوب بين رتب تلك القيم.

يتطلب حساب معامل ارتباط الرتب اعطاء رتب للقيم الاصلية للمتغيرين وتحديد مربع الفرق لكل زوج من الرتب المتقابله لهذين المتغيرين وتعويض الناتج في المعادلة التالية:

$$ر = 1 - \frac{\overset{*}{6} \sum ف^2}{ن(ن^2 - 1)}$$

مثال (7:3): فيما يلي ترتيب مجموعة من الطلاب حسب ادائهم لمهارتين مختلفتين، وذلك كما يظهر في الجدول ادناه.

---

* (6) مقدار ثابت

الطالب	الرتبه على المهاره (س)	الرتبه على المهاره (ص)	الفرق (ف)	ف^2
1	4	5	1-	1
2	8	2	6+	36
3	3	4	1-	1
4	5	1	4+	16
5	1	6	5-	25
6	7	3	4+	16
7	2	8	6-	36
8	6	7	1-	1+
			صفر	136

الحل:

1- نحسب الفرق بين الرتب لكل زوج متقابلين من رتب المتغيرين كما يظهر في العمود الرابع في الجدول اعلاه.

2- نحسب مربع الفرق بين الرتب كما هو مبين في العمود الخامس في الجدول اعلاه.

3- نعوض القيم في المعادلة التالية:

$$ر = 1 - \frac{6(ف)^2}{ن(ن^2 - 1)}$$

$$ر = 1 - \frac{6(136)}{8(64-1)}$$

$$= 1 - \frac{216}{8 \times 63}$$

$$= 1 - \frac{216}{504}$$

$$= 1 - 0.43$$

$$= 0.57$$

مثال (7:4): لديك البيانات التالية والمتعلقة بمتغيرين (س) (ص) وذلك على النحو التالي:

س (3, 15, 10 , 2, 6 , 12, 8, 5)

س (7, 16 , 4 , 5, 9 , 8 , 10 , 2)

المطلوب حساب معامل الارتباط للقيم الاصلية (معامل ارتباط بيرسون) ومعامل الارتباط لرتب القيم (معامل ارتباط سبيرمان) ثم المقارنة بينهما .

الحل:

أولا: معامل الارتباط للقيم الاصلية (ارتباط بيرسون) .

س	ص	س 2	ص 2	س × ص
5	2	025	004	10
8	10	064	100	80
12	8	144	064	96
6	9	036	081	54
2	5	004	025	10
10	4	100	016	40
15	16	225	256	240
3	7	009	049	21
61	61	607	595	551

$$ر_{س ص} = \frac{ن \sum س ص - \sum س \times \sum ص}{\sqrt{\left[ن \sum س^2 - (\sum س)^2\right]\left[ن \sum ص^2 - (\sum ص)^2\right]}}$$

$$ر_{س ص} = \frac{8(551) - (61)(61)}{\sqrt{\left[8(607) - (61)^2\right]\left[8(595) - (61)^2\right]}}$$

$$ر_{س ص} = \frac{4408 - 3721}{\sqrt{(4856 - 3721)(4760 - 3721)}}$$

$$= \frac{687}{\sqrt{(1039)\,(1135)}}$$

$$= \frac{687}{\sqrt{1179256}}$$

$$= \frac{687}{1085.94} \approx 0.63$$ تقريباً

ثانيا: معامل الارتباط لرتب القيم (معامل ارتباط سبيرمان).

رتب قيم س	رتب قيم ص	ف	ف2
3	1	2-	4
5	7	2-	4
7	5	2+	4
4	6	2-	4
1	3	2-	4
6	2	4+	16
8	8	صفر	صفر
2	4	2-	4
			40

$$ر = 1 - \frac{6\,(ف)^2}{ن\,(ن^2 - 1)}$$

$$ر = 1 - \frac{6\,(40)}{8\,(64-1)}$$

$$= 1 - \frac{240}{8\,(63)}$$

$$= 1 - \frac{240}{504}$$

$$= 1 - 0.48$$

$$= 0.52$$ تقريباً

يلاحظ ان قيمة معامل الارتباط (ارتباط بيرسون) للقيم الاصلية كانت (0.63) تقريبا في حين ان قيمة معامل الارتباط (ارتباط سبيرمان). كانت (0. 52)، وهذا مما يعني ان قيمة معامل الارتباط تختلف باختلاف طريقة حسابه.

ثالثا: معامل ارتباط فاي Phi- Coefficient:

يستخدم معامل ارتباط فاي في حالة المتغيرات النوعية التي تقع على مستوى المقياس الاسمي، بحيث يشتمل المتغير على عدة مستويات مختلفة مثل الجنس (ذكر، انثى)، ونوع المهنة (مهندس، طبيب، مزارع، عسكري . . .). . . إلخ. ومثل هذه المتغيرات لا يمكن قياسها كمياً. فالارقام التي تعطى لمستوياتها المختلفة ماهي الا لغايات تصنيفها في فئات لتسهيل عملية دراستها.

لحساب معامل الارتباط في حالة المتغيرات التصنيفيه، يتم توزيع الحالات في جدول يتألف من اربع خلايا، بحيث تأخذ كل حالة قيمتين (صفر أو 1) وذلك حسب طبيعة الظاهرتين موضع البحث. وهكذا يكون لدينا في كل خلية من الخلايا الاربع عدد من الحالات، وذلك كما هو موضح في الجدول التالي:

الجنس / الرأي حول عمل المرأة	ذكر (1)	انثى (0)	المجموع
يوافق (1)	أ	ب	أ + ب
لا يوافق (0)	ج	د	ج + د
المجموع	أ+ ج	ب+ د	

في الجدول اعلاه نلاحظ وجود متغيرين هما الجنس وله مستويان المستوى الأول ذكر وقد اعطى الرقم (1) للدلاله عليه والمستوى الثاني انثي وقد اعطى الرقم (0) للدلاله عليه. اما المتغير الثاني فهو الرأي حول عمل المرأة وله مستويان، المستوي الأول يوافق وقد اعطى الرقم (1) للدلالة عليه، اما المستوى الثاني فهو لا يوافق وقد اعطى الرقم (0) للدلالة عليه. وبالنظر إلى الخانات نلاحظ ان الرمز (أ) يشير إلى عدد الذكور الذين يوافقون على عمل المرأة، والرمز (ب) يشير إلى عدد الذكور الذين لا يوافقون على عمل المرأة، اما الرمز (ج) فهو يشير إلى عدد الاناث اللواتي يوافقن على عمل المرأة، والرمز (د) فهو يمثل عدد الاناث اللواتي لا يوافقن على عمل المرأة.

ولتحديد قيمة معامل الارتباط فاي (Ø) عادة يتم استخدام المعادلة التالية :

$$ر_Ø = \frac{أ × د - ب × ج}{\sqrt{(أ+ج)(ب+د)(ج+د)(أ+ب)}}$$

مثال (7:5) : الجدول التالي يلخص اجابات (40) طالبا من الجنسين يتعلق بموقفهم حول عمل المرأة، أحسب معامل ارتباط (Ø) لهذه البيانات :

الجنس / الرأي حول عمل المرأة	ذكر (1)	انثى (0)	المجموع
يوافق (1)	7 (أ)	15(ب)	22
لا يوافق (0)	16 (ج)	2 (د)	18
المجموع	23	17	40

الحل: $ر_Ø = \frac{16 × 15 - 2 × 7}{\sqrt{(15+7)(2+16)(2+15)(16+7)}}$

$$= \frac{240 - 14}{\sqrt{22 × 18 × 17 × 23}}$$

$$ر_Ø = \frac{226 -}{\sqrt{154836}}$$

$$= \frac{226 -}{393,50} = - 0.57$$ تقريبا

يلاحظ ان معامل الارتباط الناتج كبيرا نوعا ما رغم انه يمثل علاقة عكسيه، وهذا يعني بالطبع ان هناك علاقة قويه نوعا ما بين الجنس والاتجاه نحو عمل المرأة، حيث هناك تأييدا لعمل المرأة، وكون ان قيمة معامل الارتباط سالبه، فهذا يعني ان الاناث هن الاكثر تأييداً لعمل المرأة من الذكور.

تجدر الاشارة هنا، ان معامل الارتباط فاي (Ø) يستخدم في حالة وجود متغيرين نوعيين لكل منهما

مستويين فقط ، اما في حالة وجود أكثر من مستويين لكل المتغيرين أو احداهما عندها يستعاض عن هذه المعادلة ليتستخدم بدلا منها الاحصائي كاي2.

رابعا: معامل ارتباط بونت بايسيريال

يستخدم هذا المعامل عندما يتطلب الامر ايجاد معامل الارتباط بين متغيرين احدهما متصل والاخر نوعي يقع على مستوى القياس الاسمي ، ففي مثل هذه الحالة تستخدم صيغة معدلة من معامل ارتباط بيرسون تعرف بأسم معامل ارتباط بونت بابسيريال . فعل سبيل المثال ، قد نسعى إلى ايجاد العلاقة بين اجابات الطالب على كل فقرة (صح =1 خطأ =0) وعلامته على الامتحان الكلي ، أو العلاقة بين متوسط درجات كل من الذكور والاناث على امتحان معين . ويتم عادة حساب معامل ارتباط بونت بايسيريال من خلال استخدام المعادلة التالية :

$$ر_ب = \frac{\bar{س}_1 - \bar{س}_0}{ع_ن} \times \sqrt{ن_1\ ن_0}$$

حيث $\bar{س}_1$: متوسط درجات افراد المستوى الأول (1) من المتغير الثنائي (الفرعي)على المتغير المتصل (ص).

$\bar{س}_0$: متوسط درجات افراد المستوى الثاني(0) من المتغير الثنائي (النوعي) على المتغير المتصل (ص).

ع : الانحراف المعياري لدرجات جميع الافراد على المتغير المتصل.

ن : عدد الافراد (الحالات) الكلي .

$ن_1$: عدد افراد المستوى الأول (1).

$ن_0$: عدد افراد المستوى الثاني (0).

مثال (7:6) : لدى مدرس فصلا دراسياً يشتمل على 30 طالبا (12) ذكورا (1) أو 18 اناثا (0). اذا كان متوسط اوزان الذكور =42 كغم ومتوسط اوزان الاناث 38 كغم ، والانحراف المعياري لاوزانهم =5 . أحسب معامل الارتباط بين اوزان الاناث والذكور.

الحل : $ر_ب = \frac{\bar{س}_1 - \bar{س}_0}{ع_ن} \times \sqrt{ن_1 \times ن_0}$

$$ر = \frac{42 - 38}{5 \times 30} \times \sqrt{12 \times 18}$$

$$= \frac{4}{150} \times \sqrt{216}$$

$$= 0.27 \times 14.7$$

= 0.39 تقريباً

معامل التحديد Coefficient of Determination

يعرف معامل التحديد على انه مربع معامل الارتباط ، بحيث يستخدم هذا المعامل لتحديد مقدار التباين المشترك بين متغيرين ؛ أي تحديد مقدار التباين في احد المتغيرات التي يمكن تفيسره من خلال المتغير الاخر. فعند تربيع معامل الارتباط بين متغيرين معينين يمكن الحصول على قيمة تعرف باسم مقدار التباين المشترك أو معامل التحديد ، حيث تستخدم هذه القيمة لتفسير التباين في احد المتغيرات من خلال تحديد قيمة التباين للمتغير الاخر. فعلى سبيل المثال اذا كان معامل الارتباط بين الدخل الشهري وعدد ساعات العمل = (0.85)، فعندها يكون معامل التحديد مساويا لي $(,.85)^2$ ؛ اي (72,.) . وهذا يعني ان 72% من التباين في احد المتغيرات السابقة يمكن تفسيره من خلال تحديد التباين في المتغير الاخر والعكس صحيح. اما الجزء المتبقي من التباين فيطلق عليه اسم التباين العشوائي ، ومثل هذا التباين لا يمكن تفسيره من خلال التباين في احد المتغيرات لانه قد يعزى إلى عوامل اخرى مثل عوامل الصدفه أو بسبب وجود متغيرات اخرى تؤثر في المتغيرات موضع البحث.

مثال (7:7) : اذا كان معامل الارتباط بين كمية السماد وحجم الانتاج النباتي = 0. 92 فكم من التباين في احد هذين المتغيرين يمكن تفسيره من خلال معرفة التباين في المتغير الاخر.

الحل: نحسب معامل التحديد ، وهو بمثابة مربع معامل الارتباط.

اذا معامل التحديد = $(ر)^2 = (0.92)^2$.

= 0.85 تقريبا.

وهكذا نجد ان 85% من التباين في احد المتغيرات السابقة يمكن تفسيره من خلال تحديد التباين في المتغير الاخر، اما الجزء المتبقي من التباين (0.15) فقد يعود إلى عوامل أو متغيرات اخرى.

الانحدار والتنبوء Regression and Prediction

لقد عرفنا سابقا ان معامل الارتباط هو مؤشر احصائي يستفاد منه في الكشف عن العلاقات بين المتغيرات. فقد يكشف هذا المتغير عن طبيعة العلاقة من حيث كونها سالبة أو موجبة أو صفرية (محايدة)، وفيما اذا كانت هذه العلاقة قوية أو ضعيفة أو متوسطة. هذا ويمكن ان يكون معامل الارتباط مؤشرا دقيقا للتنبوء بقيمة متغير ما في ضوء تقديرها وفقا لقيمة متغير اخر يرتبط به. فعلى سبيل المثال، اذا كشفت النتائج وجود علاقة معقولة بين عدد ساعات العمل وكمية الانتاج، فعندها بالتالي يمكن تقدير كمية الانتاج من خلال تحديد ساعات العمل؛ أي تقدير كمية الانتاج عند كل مستوى من مستويات ساعات العمل وذلك من خلال معادلة تعرف باسم معادلة الانحدار أو التنبوء.

يوجد نوعين من معادلات الانحدار أو التنبوء تبعا لعدد المتغيرات التي تشتمل عليها وهي:

1- معادلة الانحدار الخطي البسيط Simple linear regression، نستخدم هذه المعادلة في حالة وجود علاقة خطية بين متغيرين وذلك من خلال استخدام معادلة الخط المستقيم.

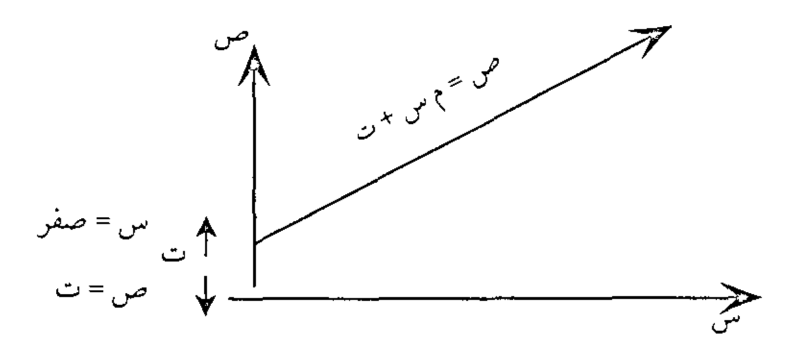
2- معادلة الانحدار الخطي المتعدد Mulitple linear regression : وتستخدم هذه المعادلة عندما يوجد لدينا اكثر من ثلاثة متغيرات ويراد تقدير قيمة احدهما من خلال المعلومات المتوفرة عن المتغيرات الاخرى.

- الانحدار الخطي البسيط:

قبل الحديث عن فكرة الانحدار الخطي البسيط فانه من المستحسن التعرف على مفهوم الارتباط الخطي أو العلاقة الخطية. تشير العلاقة الخطية إلى وجود علاقة تامة بين متغيرين (س، ص). بحيث يمكن تحديد قيمة (ص) بدقة تامة من خلال قيمة (س) وذلك وفق المعادلة التالية:

ص = م $_{س}$ + ت، حيث (م & ت) هما ثابتان.

إذ ان (م) يمثل ميل الخط المستقيم على المحور الافقي (السيني) وهو يعكس مقدار التغير في المتغير (ص) عند تغير المتغير (س) بمقدار وحدة واحدة. اما (ت) فهو يمثل نقطة تقاطع الخط المستقيم مع المحور العمودي (ص) عندما تكون قيمة المتغير (س) = صفر؛ اي هو ثابت الخط المستقيم أو قيمة المتغير (ص) عندما تكون قيمة المتغير (س) تساوي صفرا، وذلك كما هو موضح في الشكل التالي:



فعلى سبيل المثال، اذا كانت العلاقة بين المتغيرين (س، ص) تامه وكانت قيمة كل من (م،)و (ت) معروفة لدينا بحيث تساو ي 2و 1 على التوالي، عندها يمكن تحديد قيم المتغير (ص) بدقة متناهية عند اي قيمة من قيم المتغير (س) وذلك على النحو التالي:

ص = م س × ت

فاذا كانت قيمة س = 1 ، فعندها فان قيمة ص = 2 × 1 +1= 3، اما اذا كانت قيمة س = 2 ، فان قيمة ص = 2 × 2 +1 = 5، في حين اذا كانت قيمة س = 3 فعندها فان قيمة ص= 2 ×3 +1= 7. لاحظ ان قيم المتغير (ص) تزداد بمقدار وحدتين عندما يزداد المتغير (س) بمقدار واحده، ففي مثل هذه الحالة فان العلاقة بينهما تسير على نحو خطي منتظم وفق معادلة الخط المستقيم.

إن مثل هذه العلاقة يمكن ايجادها في حالة العديد من المتغيرات مثل العلاقة بين الكيلو مترات والاميال، وذلك وفق المعادل التالية:

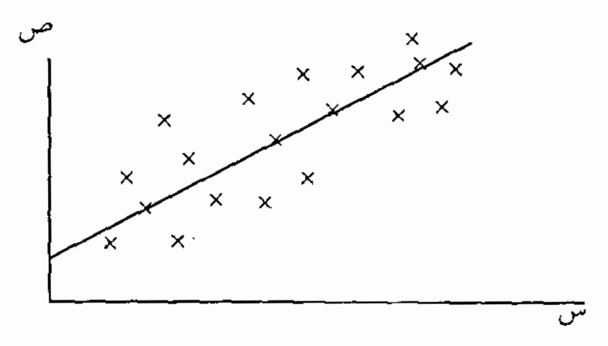
الميل = $\frac{5}{8}$ × الكيلومترات.

فاذا كانت المسافة المقطوعة ص = 30 كيلو متر، فان هذه المسافة تناظر اميالا مقدارها $\frac{5}{8}$ × 30 = 18.75 ميل تقريبا. كذلك يمكن القول بان العلاقة بين الدرجات الحرارية المئوية والفهرنهايتيه هي من النوع الخطي وفق المعادلة التالية:

الدرجة الفهرنهايتيه = $\frac{9}{5}$× الدرجة المئوية + 32. فاذا كانت الدرجة المئوية = 35 فعندها

فان الدرجة الفهرنهايتيه = $\frac{9}{5}$ × (35) + 32 = 63 + 32 = 95.

ولكن في حالة المتغيرات النفسية والتربوية والإنسانية مثل الذكاء والتحصيل والدافعية وغيرها من المتغيرات الاخرى، فان من الصعب ايجاد علاقة تامة بين مثل هذه المتغيرات، الامر الذي يجعل من المستحيل تمثيلها بيانيا من خلال خط مستقيم يعبر بصدق عن جميع الحالات. ففي مثل حالة هذه المتغيرات نلجأ إلى تمثيلها بيانيا من خلال خط مستقيم يعرف باسم خط افضل تمثيل (Line of best fit) وذلك باستخدام فكره القيمة الصغرى للمربعات (Least Square Method)، اي من خلال رسم خط انتشار يكون فيه مربع انحرافات القيم عن هذا الخط اقل ما يمكن؛ وهو الخط الذي يمر في اكبر عدد ممكن من النقاط المثله لقيم المتغيرين، او انه الخط الذي يتجمع حوله اكبر عدد ممكن من النقاط، وذلك كما هو موضح في الشكل التالي:



وتجدر الاشارة هنا، ان طريقة المربعات الصغرى تتيح لنا رسم افضل خط لانتشار القيم، وهو ما يعرف بخط الانحدار بحيث نلاحظ ان مجموع انحرافات القيم عن هذا الخط تساوي صفرا. ان اقتراب القيم من هذا الخط يعد مؤشرا إلى قوة العلاقة بين المتغيرين. مما يتيح لنا امكانية التنبوء بالمتغير التابع على نحو دقيق من خلال قيم المتغير المستقل.

عموما، في حالة عدم وجود علاقة تامه بين المتغيرين يتم الاعتماد على التمثيل الرياضي عوضا عن التمثيل البياني، وذلك من خلال معادلة تعرف بمعادلة خط الانحدار وذلك على النحو الاتي:

$\hat{ص}$: م $ص_{س}$ × س + $ت_{ص س}$

حيث: $\hat{ص}$ هي قيمة المتغير (ص) المتنبأ بها

م $_{ص س}$:ميل الانحدار.

ت $_{ص س}$: ثابت الانحدار وهو عبارة عن قيمة المتغير (ص) عندما تكون قيمة المتغير (س) = صفرا

هذا ويمكن حساب قيمة كل من $م_{ص س}$ و $ت_{ص س}$ من خلال المعادلات التالية :

$$م_{ص س} = \frac{ن \sum س ص - (\sum س)(\sum ص)}{ن \sum س^2 - (\sum س)^2}$$

أو

$$م_{ص س} = ر_{س ص} \times \frac{ع_{ص}}{ع_{س}}$$

حيث : $ر_{ص س}$ يمثل قيمة معامل الارتباط بين المتغير س والمتغير ص .

$ع_{س}$: الانحراف المعياري لقيم المتغير (ص) .

$ع_{س}$: الانحراف المعياري لقيم المتغير (س) .

$$ت_{ص س} = \bar{ص} - م_{ص س} \times \bar{س}.$$

حيث : $\bar{ص}$: هو المتوسط الحسابي لقيم (ص) .

$\bar{س}$: هو المتوسط الحسابي لقيم (س) ، $م_{ص س}$: ميل خط الانتشار

مثال (7:8) : الجدول ادناه يبين قيم كل من المتغير المستقل (س) والمتغير التابع (ص) ، المطلوب حساب قيمة كل من $م_{ص س}$ و $ت_{ص س}$ ، ثم تقدير قيم المتغير (ص) من المتغير (س) .

س 2	س ص	ص	س
01	04	4	1
25	35	7	5
16	24	6	4
49	56	8	7
09	15	5	3
100	134	30	20

الحل: لحساب $م_{ص س}$ نتبع الخطوات التالية :

1- نجد ناتج مجموع قيم المتغير (س) والمتغير (ص) كما يظهر في اسفل العمودين الأول والثاني في الجدول اعلاه .

2- نجد مجموع ناتج ضرب قيم المتغير (س) بقيم المتغير (ص)، كما هو مبين في العمود الثالث في الجدول اعلاه.

3- نحسب مجموع مربع قيم المتغير (س) وذلك كما يظهر في اسفل العمود الرابع في الجدول اعلاه.

4- نحسب مربع مجموع قيم المتغير (س)؛ اي $(\sum س)^2 = (20)^2 = 400$.

5- نعوض القيم في المعادلة التالية لنحصل على قيمة ($م_{ص س}$).

$$م_{ص س} = \frac{ن \sum ص س - (\sum س)(\sum ص)}{ن \sum س^2 - (\sum س)^2}$$

$$م_{ص س} = \frac{5(134) - (20)(30)}{5(100) - 400}$$

$$م_{ص س} = \frac{670 - 600}{500 - 400} = \frac{70}{100} = 0.70$$

ب: لحساب $ت_{ص س}$ نتبع الخطوات التالية:

1- نحسب المتوسط الحسابي لقيم المتغير (ص).

$$\bar{ص} = \frac{\sum ص}{ن} = \frac{30}{5} = 6$$

2- نحسب المتوسط الحسابي لقيم المتغير (س).

$$\bar{س} = \frac{\sum س}{ن} = \frac{20}{5} = 4$$

3- نعوض القيم بالمعادلة التالية:

$$ت_{ص س} = \bar{ص} - م_{ص س} \times \bar{س}$$

$ت_{ص س} = 6 - 0.70 \times 4$.

$= 6 - 2.8$

$= 3.2$

وعليه فان معادلة الانحدار لتقدير قيم المتغير (ص) من المتغير (س) تصبح على النحو الاتي:

$\hat{ص} = م_{ص س} \times س + ت_{ص س}$

$= .070$ (س) $+ 3.2$

مثال: إذا كانت قيمة (س) = (4)، فكم تكون قيمة (ص) المتنبأ بها

الحل:

$\hat{ص} = 0.70 \times 4 + 3.2$

$= 2.8 + 3.2 = 6$

مثال: إذا كانت قيمة (س) = (7)، فكم تكون قيمة (ص) المتنبأ بها.

$\hat{ص} = 0.70 \times 4 + 3.2$

$= 4.9 + 3.2 = 8.1$

لاحظ قيم (ص) المتنبأ بها وقارنها مع تلك الموجودة في الجدول، نلاحظ أنها متطابقة معها، مما يعني أن هذه المعادلة تتنبأ بشكل جيد بقيم المتغير (ص) من قيم المتغير (س).

في المثال السابق يمكن حساب م ص س بطريقة أخرى وذلك من خلال حساب معامل الارتباط بين قيم المتغير (س) وقيم المتغير (ص) وحساب الانحراف المعياري لقيم (س) وقيم (ص) وذلك وفق المعادلة التالية:

$$م_{ص س} = ر_{ص س} \times \frac{ع_{ص}}{ع_{س}}$$

الحل:

س	ص	$س^2$	$ص^2$	س ص
1	4	1	16	04
5	7	25	49	35
4	6	16	36	24
7	8	49	64	56
3	5	09	25	15
20	30	100	190	134

$$ر_{ص س} = \frac{ن \sum_{س ص} - (\sum س)(\sum ص)}{\sqrt{\left[ن \sum س^2 - (\sum س)^2\right]\left[ن \sum ص^2 - (\sum س)^2\right]}}$$

$$= \frac{5(134) - 20 \times 30}{\sqrt{\left[5(100) - (20)^2\right]\left[5(190) - (30)^2\right]}}$$

$$= \frac{670 - 600}{\sqrt{[500 - 400][950 - 900]}}$$

$$= \frac{70}{\sqrt{100 \times 50}} = \frac{70}{\sqrt{5000}} = \frac{70}{70.7} \approx 0.99$$

$$ع_{ص} = \sqrt{\frac{\sum ص^2 - \frac{(\sum ص)^2}{ن}}{ن - 1}}$$

$$= \sqrt{\frac{190 - \frac{900}{5}}{5 - 1}} = \sqrt{\frac{190 - 180}{4}}$$

$$= \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58$$

$$ع_{س} = \sqrt{\frac{\sum س^2 - \frac{(\sum س)^2}{ن}}{ن - 1}}$$

$$= \sqrt{\frac{100 - \frac{400}{5}}{5 - 1}} = \sqrt{\frac{100 - 80}{4}} = \sqrt{5} = 2.24$$

وعليه فان $م_{صس} = ر_{صس} \times \frac{ع_{ص}}{ع_{س}}$

$0.99 \times \frac{1.58}{2.24} \approx 0.70$ تقريبا وهي نفس القيمة التي تم الحصول عليها في المعادلة السابقة .

تستخدم المعادلات السابقة لتقدير أو التنبوء بقيم المتغير التابع (ص) من قيم المتغير المستقل (س)، ولكن السؤال الذي يطرح نفسه هنا هو: هل بالامكان التنبوء بقيم المتغير المستقل من خلال معرفة قيم المتغير التابع ؛ وبالطبع ان الاجابة عن هذا السؤال هي بالايجاب، إذ يمكن تقدير قيم المتغير المستقل من خلال قيم المتغير التابع وفقا للمعادلات التالية :

$\hat{س} = م_{س ص} \times ص + ت_{س ص}$ ← معادلة التنبوء (الانحدار)

$$م_{س ص} = \frac{ن \sum س ص - (\sum س)(\sum ص)}{ن \sum ص^2 - (\sum ص)^2}$$ ← معادلة تحديد الميل

الحل:

س	ص	$س^2$	$ص^2$	س ص
1	4	1	16	04
5	7	25	49	35
4	6	16	36	24
7	8	49	64	56
3	5	09	25	15
20	30	100	190	134

$$ر_{\text{ص س}} = \frac{\text{ن}\sum \text{س ص} - (\sum \text{س})(\sum \text{ص})}{\sqrt{\left[\text{ن}\sum \text{س}^2 - (\sum \text{س})^2\right]\left[\text{ن}\sum \text{ص}^2 - (\sum \text{س})^2\right]}}$$

$$= \frac{5\,(134) - 20 \times 30}{\sqrt{\left[5\,(100) - (20)^2\right]\left[5\,(190) - (30)^2\right]}}$$

$$= \frac{670 - 600}{\sqrt{\left[500 - 400\right]\left[950 - 900\right]}}$$

$$= \frac{70}{\sqrt{100 \times 50}} = \frac{70}{\sqrt{5000}} = \frac{70}{70.7} \approx 0.99$$

$$ع_{\text{ص}} = \sqrt{\frac{\sum \text{ص}^2 - \frac{(\sum \text{ص})^2}{\text{ن}}}{\text{ن} - 1}}$$

$$= \sqrt{\frac{190 - \frac{900}{5}}{5 - 1}} = \sqrt{\frac{190 - 180}{4}}$$

$$= \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58$$

$$ع_{س} = \sqrt{\frac{\sum س^2 - \frac{(\sum س)^2}{ن}}{ن - 1}}$$

$$= \sqrt{\frac{100 - \frac{400}{5}}{5 - 1}} = \sqrt{\frac{100 - 80}{4}} = \sqrt{5} = 2.24$$

وعليه فان $م_{ص س} = ر_{ص س} \times \frac{ع_{ص}}{ع_{س}}$

$0.99 \times \frac{1.58}{2.24} \approx 0.70$ تقريبا وهي نفس القيمة التي تم الحصول عليها في المعادلة السابقة .

تستخدم المعادلات السابقة لتقدير أو التنبوء بقيم المتغير التابع (ص) من قيم المتغير المستقل (س)، ولكن السؤال الذي يطرح نفسه هنا هو: هل بالامكان التنبوء بقيم المتغير المستقل من خلال معرفة قيم المتغير التابع ؛ وبالطبع ان الاجابة عن هذا السؤال هي بالايجاب، إذ يمكن تقدير قيم المتغير المستقل من خلال قيم المتغير التابع وفقا للمعادلات التالية :

$\hat{س} = م_{س ص} \times ص + ت_{س ص}$ ← معادلة التنبوء (الانحدار)

$م_{س ص} = \frac{ن \sum س ص - (\sum س)(\sum ص)}{ن \sum ص^2 - (\sum ص)^2}$ ← معادلة تحديد الميل

$$أو = ر_{س ص} \times \frac{ع_{س}}{ع_{ص}}$$

$$ت_{س ص} = \bar{س} - م_{س ص} \times \bar{ص}$$ ← معادلة ثابت التقاطع

مثال (7:9) : في الجدول أدناه بيانات تتعلق بنتائج اختيارات 5 طلاب على مادتين التربية الإسلامية والاجتماعيات .

س	ص	س ص	$س^2$
36	43	1548	1296
18	36	648	324
24	50	1200	576
33	56	1848	1089
24	45	1080	576
135	230	6324	3851

المطلوب :

أ) ايجاد معادلة انحدار المتغير (ص) التحصيل في التربية الإسلامية على المتغير (س) التحصيل في الاجتماعيات ، أي التنبؤ بقيم (ص) من قيم المتغير (س) .

ب) ايجاد معادلة انحدار المتغير (س) التحصيل في الاجتماعيات (س) على المتغير (ص) التحصيل في التربية الإسلامية ، أي التنبؤ بقيم (س) من قيم المتغير (ص) .

أ. **الحل:**

(الفرع الأول من السؤال) .

1 - نحسب قيمة م $_{ص س}$

$$م_{ص س} = \frac{ن \sum س ص - (\sum س)(\sum ص)}{ن \sum س^2 - (\sum س)^2}$$

$$= \frac{5(6324) - (135)(230)}{5(3851) - (135)^2}$$

$$= \frac{31620 - 31050}{19255 - 18225} = \frac{570}{955} = 0.58$$

إذاً قيمة م ص س = 0.58، وهذا يعني أن التحصيل في مادة التربية الإسلامية (ص) يزداد بمقدار (0.58) نقطة كلما ازداد التحصيل في الاجتماعيات بمقدار وحدة واحدة.

2- نحسب قيمة ت $_{ص س}$.

$$ت_{ص س} = \bar{ص} - م_{ص س} \times \bar{س}$$

$$\bar{ص} = \frac{\sum ص}{ن} = \frac{230}{5} = 46$$

$$\bar{س} = \frac{\sum س}{ن} = \frac{135}{5} = 27$$

إذاً ت $_{ص س}$ = (46) - (0.58) × 27.

$$46 - 15.66 = 30.44$$

هذا يعني أن التحصيل في التربية الإسلامية يساوي (30.44) عندما يكون مقدار التحصيل في الاجتماعيات يساوي صفراً. للتنبؤ بأي قيمة من قيم المتغير التابع (ص) وهو التحصيل في مادة التربية الإسلامية من خلال قيم المتغير المستقل (س) التحصيل في الاجتماعيات، يمكن استخدام المعادلة التالية :

$$\hat{ص} = م_{ص س} \times س + ت_{ص س}$$

فعلى سبيل المثال إذا كانت قيمة س = 47، فإن قيمة (ص) المتنبأ بها هي:

$0.58 \times 45 + 30.44 = 56.54$ تقريباً

ب. الحل: (الفرع الثاني من السؤال).

للتنبؤ بقيم المتغير المستقل (س) التحصيل في الاجتماعيات من قيم المتغير التابع (ص) التحصيل في التربية الإسلامية تستخدم المعادلة التالية:

$$\hat{س} = م_{س ص} \times ص + ت_{س ص}$$

س	ص	س ص	$س^2$
36	43	1548	1849
18	36	648	1296
24	50	1200	2500
33	56	1848	3136
24	45	1080	2025
135	230	6324	10806

$$م_{س ص} = \frac{ن \sum س ص - \left(\sum س\right)\left(\sum ص\right)}{ن \sum ص^2 - \left(\sum ص\right)^2}$$

$$= \frac{5\,(6324) - (135)(230)}{5\,(10806) - (230)^2}$$

$$= \frac{31620 - 31050}{54030 - 52900} = \frac{570}{1130} = 0.50$$

إذاً $م_{س ص} = 0.50$، وهذا يعني أن التحصيل في مادة الاجتماعيات يزداد بمقدار (0.50) نقطة عندما يزداد التحصيل في التربية الإسلامية بمقدار وحدة واحدة.

$$ت_{س ص} = \bar{س} - م_{س ص} \times \bar{ص}$$

$$\bar{س} = 27$$

$$\bar{ص} = 46$$

إذا $ت_{س ص} = (27) - (0.50)(46)$

$= 27 - 23 = 4$

وهذا يعني أن التحصيل في الاجتماعيات يساوي (4) عندما يكون مقدار التحصيل في التربية الإسلامية يساوي صفراً. وللتنبؤ بأي قيمة من قيم المتغير المستقل (التحصيل في الاجتماعيات (س) من قيم المتغير التابع التحصيل في التربية الإسلامية (ص)، يمكن استخدام المعادلة التالية:

$$\hat{س} = م_{س ص} \times ص + ت_{س ص}$$

فإذا كانت قيمة ص = 50، فإن قيمة (س) المتنبأ بها تساوي

$= 0.50 \times 50 + 4$

$= 25 + 4 = 29$

- **خطأ التقدير** Error of Estimation

يعرّف خطأ التقدير على أنه انحراف القيم المتنبأ بها للمتغير التابع عن القيم الأصلية لذلك المتغير، وهو يعرف بخطأ التنبؤ بحيث يرمز له بالرمز (خ). ويمكن أن يشار لهذا الخطأ رياضياً من خلال المعادلة الرياضية التالية:

$خ = ص - \hat{ص}$ حيث ص تمثل القيمة الاصلية، في حين آن $\hat{ص}$ تمثل القيمة المتنبأ بها.

إن مثل هذا الخطأ يمكن أن يكون قليلاً عندما تكون العلاقة الارتباطية قوية بين المتغير التابع (ص) والمتغير المستقل (س)، في حين أن قيمته تزداد في حالة كون العلاقة بينهما ضعيفة. وعموماً فإنه في حالة عدم وجود علاقة تامة بين المتغيرين فإن قيم التنبؤ بأي متغير من قيم المتغير الأخر تنزع إلى الانحدار نحو المتوسط أو ما يسمى بخط العلاقة المتوسطة.

• **الخطأ المعياري لتقدير Standard Error of Estimation**

يعرف هذا الخطأ باسم الخطأ المعياري للتنبؤ "Standard Error of Prediction" وهو مؤشر لمدى دقة معادلة الانحدار بالتنبؤ بالعلاقة بين متغيرين. فلما كان خط الانحدار لا يمر بجميع النقاط الذي يمثلها، فمن الطبيعي أن نجد أن القيم المتنبأ بها والتي يتم حسابها من خلال معادلة الانحدار لا تكون متطابقة تماماً مع القيم الأصلية للمتغير موضع البحث. ومثل هذا الاختلاف أو الفرق يزداد كلما كانت العلاقة بين المتغيرين التابع والمستقل ضعيفة.

ولقياس دقة التنبؤ، ودرجة التوافق بين القيم الأصلية للمتغير التابع وقيمته المحسوبه من معادلة الانحدار، نلجأ إلى استخدام ما يسمى بالخطأ المعياري للتنبؤ، وهو عبارة عن الانحراف المعياري للأخطاء أو الفروق بين القيم الأصلية والقيم المتنبأ بها للمتغير التابع موضوع البحث. وعادة يتم حساب الانحراف المعياري لاخطاء التنبؤ من خلال إحدى المعادلتين التاليتين:

1- الخطأ المعياري للتنبوء خ ع $_{ص س}$ $= \sqrt{\frac{\sum خ^2}{ن-1}}$

حيث $\sum خ^2$: هي مجموع مربعات الفروق بين القيم الأصلية والقيم المتنبأ بها للمتغير التابع.

ن: هي عدد الحالات.

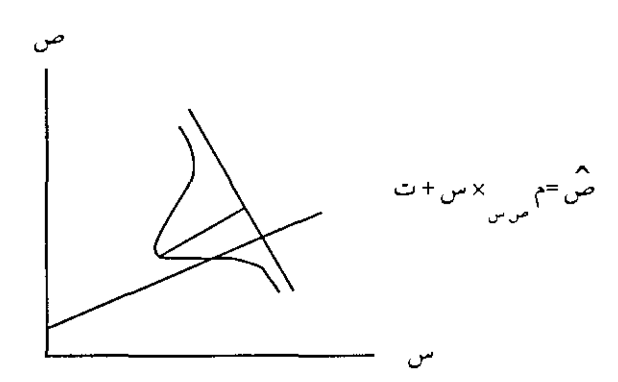
2- الخطأ المعياري للتنبوء خ $_{ص س}$ $=$ ع $_{ص}$ $\sqrt{1 - ر^2}$

حيث ع $_{ص}$: هو الانحراف المعياري لقيم المتغير التابع الأصلية.

$ر^2$: مربع معامل الارتباط بين المتغير التابع (ص) والمتغير المستقل (س).

يلاحظ أن أخطاء التنبؤ أو ما يعرف بأخطاء التقدير تتوزع توزيعاً طبيعياً بمتوسط حسابي مقداره صفر وبانحراف معياري لقيم المتغير التابع (ص) عند كل قيمة من قيم المتغير المستقل (س)، حيث أن قيمة هذا الخطأ تقترب من الصفر كلما اقتربت قيمة معامل الارتباط بين المتغير التابع والمتغير المستقل من الواحد

صحيح ، أي عندما تتجه العلاقة بينهما لتكون تامة . وهذا يتضمن افتراضاً مفاده أن أخطاء التقدير يجب أن تكون عشوائية ومستقلة عن بعضها البعض ، إذ أن الخطأ المرتبط بتقدير قيمة معينة من المتغير التابع (ص) لا تؤثر أو تتأثر بالأخطاء المرتبطة بالقيم الأخرى من ذلك المتغير. وهذا يعني أن أخطاء التنبؤ تأخذ شكل التوزيع الطبيعي بمتوسط مقداره صفر وانحراف معياري يساوي الانحراف المعياري لقيم المتغير التابع الأصلية وذلك كما يظهر بالشكل التالي :



مثال (7:10) : احسب الخطأ المعياري للتنبؤ بقيم المتغير (ص) من المتغير (س) في الجدول أدناه؟

س	ص	$ص^2$	$س^2$	س ص
2	6	36	04	12
5	3	09	25	15
3	8	64	09	24
6	9	81	36	54
4	4	16	16	16
20	30	206	90	120

الحل : يمكن حساب الخطأ المعياري للتنبؤ من خلال المعادلة التالية :

$$ع خ_{ص س} = ع_{ص} \sqrt{1 - ر^2}$$

ويتطلب ذلك إجراء الخطوات التالية :

(1) حساب الانحراف المعياري لقيم (ص) الأصلية .

(2) حساب معامل الارتباط بين قيم (ص) وقيم (س) ثم ايجاد مربع معامل الارتباط بينهما .

(3) تعويض القيم في المعادلة أعلاه .

$$\text{ع}_{\text{ص}} = \sqrt{\frac{\sum \text{ص}^2 - \frac{\left(\sum \text{ص}\right)^2}{\text{ن}}}{\text{ن} - 1}} = \sqrt{\frac{206 - \frac{(30)^2}{5}}{5 - 1}}$$

$$= \sqrt{\frac{206 - \frac{900}{5}}{4}}$$

$$= \sqrt{\frac{206 - 180}{4}} = \sqrt{\frac{26}{4}} = 2.55$$

$$\text{ر}_{\text{س ص}} = \frac{\text{ن} \sum \text{س ص} - \left(\sum \text{س}\right)\left(\sum \text{ص}\right)}{\left[\text{ن} \sum \text{س}^2 - \left(\sum \text{س}\right)^2\right]\left[\text{ن} \sum \text{ص}^2 - \left(\sum \text{ص}\right)^2\right]}$$

$$= \frac{5\,(121) - (20)(30)}{\left[5\,(90) - (20)^2\right]\left[5\,(206) - (30)^2\right]}$$

$$= \frac{605 - 600}{\left[450 - 400\right]\left[1030 - 900\right]}$$

$$= \frac{5}{\sqrt{(50)\,(130)}} = \frac{5}{\sqrt{6500}} = \frac{5}{80.62}$$

إذا $ر_{ص س} = 0.062$

$$\left(ر_{ص س}\right)^2 = \left(0.062\right)^2 = 0.003844$$

$$ع خ_{ص س} = (2.55)\sqrt{1-0.003844}$$

$$= (2.55)\sqrt{0.996}$$

$$= 2.55\ (0.998) = 2.545$$

كما يمكن حساب هذا الخطأ المعياري من خلال المعادلة التالية :

$ع خ_{ص س} = \sqrt{\frac{\sum خ^2}{ن-1}}$ وذلك على النحو التالي :

س	ص	$س^2$	س ص	$\hat{ص} = 0.10(س) + 5.6$	$خ = ص - \hat{ص}$	$خ^2 = (ص - \hat{ص})^2$
2	6	04	12	5.8	,.20+	,.4
5	3	25	15	6.1	3.10-	9.61
3	8	09	24	5.9	2.10+	4.41
6	9	36	54	6.2	2.8+	7.84
4	4	16	16	6	2.00-	4.00
20	30	901	121	30	صفر	25.9

1) $خ = ص - \hat{ص}$

2) $\hat{ص} = م_{ص س} \times س + ت_{ص س}$

$$م_{ص س} = \frac{ن \sum س ص - (\sum س)(\sum ص)}{ن \sum س^2 - (\sum س)^2}$$

$$= \frac{5 (121) - (20)(30)}{5 (90) - 400}$$

$$= \frac{605 - 600}{50} = \frac{5}{50} = 0.10$$

3) $ت_{ص س} = \bar{ص} - م_{ص س} \times \bar{س}$

$$\bar{ص} = \frac{\sum ص}{ن} = \frac{30}{5} = 6$$

$$\bar{س} = \frac{\sum س}{ن} = \frac{20}{5} = 4$$

اذا $ت_{ص س} = 6 - (0.10)(4)$.

$= 6 - 0.4$

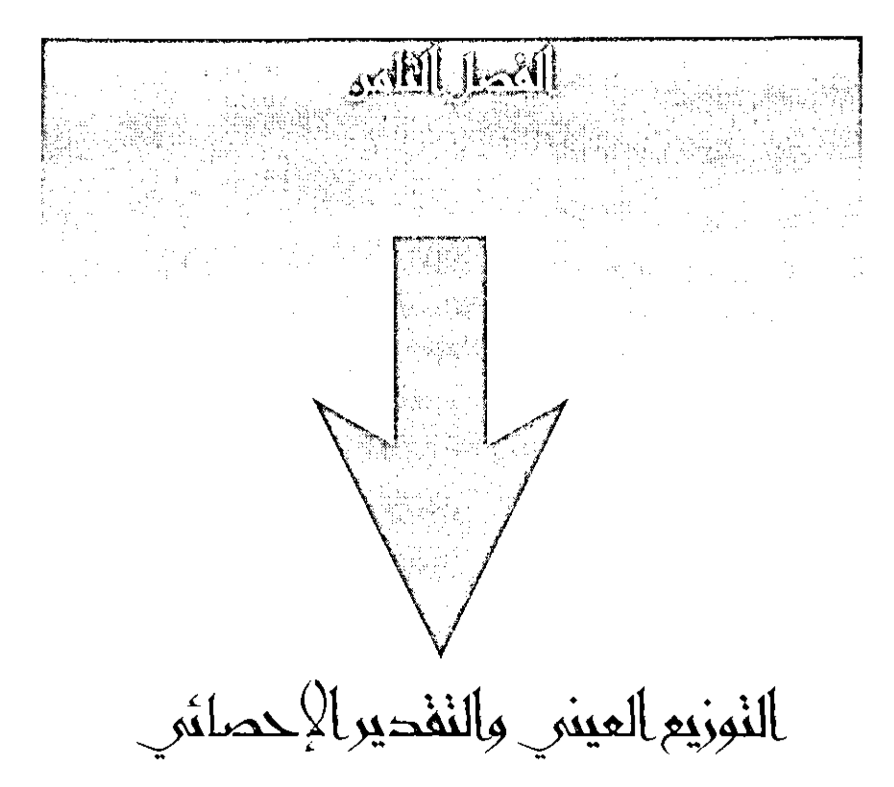
$= 5.6$

وعليه فان معادلة الانحدار أو التنبوء تصبح على النحو التالي :

ص= 0.10 (س) + 5.6 اعتمادا على هذه المعادلة يمكن ايجاد قيم (ص) المتبنأ بها وذلك كما يظهر في العمود الخامس في الجدول اعلاه . ثم العمل على ايجاد الفروق بين قيم (ص) الاصلية وقيم (ص) المتنبأ بها كما يظهر في العمود السادس، ثم ايجاد مربع هذه الفروق لتمثل ($خ^2$)، كما هو واضح في العمود السابع في الجدول اعلاه. وبتعويض القيم في المعادلة التالية نحصل على قيمة الخطأ المعياري للتنبوء.

$$ع خ_{ص س} = \sqrt{\frac{\sum خ^2}{ن-1}} = \sqrt{\frac{25.9}{4}} = 2.545$$

وهو نفس الجواب الذي تم الوصول إليه في المعادلة السابقة .



التوزيع العيني والتقدير الإحصائي

Sampling distribution and Statistical Estimation

تمهيد

لقـد علمنا سـابقاً أن العـينة هي مجـموعـة جزئية يتم سحـبها من مجـتمع إحصـائي معين لدراسة خصائصها وحساب بعض القيم الإحصائية من خلال بياناتها والتي تستخدم كمؤشر للاستدلال على خصائص المجتمع الذي أخذت منه. وتكمن الفلسفة وراء دراسة العينات بدلاً من المجتمعات الإحصائية في كون أن المجتمعات الإحصائية في الكثير من الأحيان تكون كبيرة جداً أو غير محددة، ومن هنا بات من الضروري اللجوء إلى استخدام المفهوم العيني في الإحصاء. تسمى القيم الإحصائية التي يتم حسابها من خلال بيانات العينة مثل المتوسط والوسيط والانحراف المعياري ومعامل الارتباط وغيرها بالإحصائيات "Statistics"، ومثل هذه الإحصائيات تستخدم لتقدير الإحصائيات المناظرة لها للمجتمع ككل بحيث يطلق على إحصائيات المجتمع اسم المعالم "Parameters".

تعتمد دقة تقدير معالم المجتمع من خلال بيانات (إحصائيات) العينة على مدى تمثيل عناصر العينة للمجتمع الإحصائي الذي أخذت منه، بحيث تقترب قيم إحصائيات العينة من المعالم الحقيقية للمجتمع الإحصائي في حالة كون حجم العينة كبيراً وفي كونها تمثل ذلك المجتمع. هذا وتعد العينة ممثلة للمجتمع عندما تتوزع فيها الخصائص بنفس النسب أو بنسب متقاربة لتلك الموجودة في المجتمع الذي أخذت منه.

○ التوزيع العيني (نظرية النهاية المركزية):

تشير نتائج الدراسات الإحصائية إلى أنه في حالة سحب عدد كبير من العينات من حجم ثابت من مجتمع إحصائي معين، تتوزع فيه الخصائص توزيعاً طبيعياً، فإن توزيع المتوسطات لهذه العينات ينزع ليأخذ شكل التوزيع الطبيعي بمتوسط حسابي يساوي المتوسط الحسابي للمجتمع الإحصائي وبانحراف معياري يساوي ناتج قسمة الانحراف المعياري للمجتمع على الجذر التربيعي لحجم العينة $\left(\frac{\text{ن}}{\sqrt{\text{ن}}}\right)$.

فعندما يتم أخذ عـدداً كبيـراً من العينات التي لها نفس الحجم ، فإن العلاقة بين معلمات المجتمع وإحصائيات تلك العينات يمكن تحديدها على النحو الآتي :

1- المتوسط الحسابي للمجتمع (u) هو بمثابة المتوسط العام لمتوسطات تلك العينات ، أي مجموع متوسطات العينات مقسوماً على عدد العينات وفق المعادلة التالية :

$$\text{متوسط المجتمع} = \frac{\sum \text{متوسطات العينات}}{\text{عدد العينات}}$$

2- التباين للمجتمع– المتوسط العام لتباين العينات ، أي مجموع التباينات لكل العينات مقسوماً على عدد العينات ، وفق المعادلة التالية :

$$\text{تباين المجتمع} = \frac{\sum \text{تباينات العينات}}{\text{عدد العينات}}$$

ولكن تجدر الإشارة هنا ، أنه في حالة أخذ عدد كبير من العينات من مجتمع إحصائي فمن الطبيعي أن نلاحظ وجود تباين في قيم المتوسطات الحسابية لتلك العينات ، إذ نجد أن بعضها يكون مساوياً في قيمته لقيمة متوسط المجتمع والبعض الآخر قد يكون أكبر أو أصغر في قيمته من قيمة متوسط المجتمع . وهكذا نجد أن هناك تشتتاً في قيم المتوسطات الحسابية للعينات حول المتوسط الحسابي للمجتمع ، ومثل هذا الانحراف أو التشتت يطلق عليه اسم الخطأ المعياري بحيث يمثل رياضياً وفق المعادلة التالية :

$$\text{خ ع} = \frac{\text{ع}}{\sqrt{\text{ن}}}$$

حيث خ ع : الخطأ المعياري

ع : الانحراف المعياري للمجتمع

ن : حجم العينة

وانطلاقاً من ذلك ، واعتماداً على نظرية النهاية المركزية ، فإن متوسطات العينات تنزع لتأخذ شكل التوزيع الطبيعي على نحو مماثل لتوزيع المشاهدات في المجتمع الأصلي الذي أخذت منه هذه العينات

بمتوسط حسابي يساوي المتوسط الحسابي للمجتمع الإحصائي وبانحراف معياري يسمى بالخطأ المعياري، وهو عبارة عن ناتج قسمة الانحراف المعياري للمجتمع الأصلي على عدد أفراد العينة.

إن قيمة الخطأ المعياري تقل مع زيادة حجم العينة، وهذا يعني أنه بزيادة حجم العينة فإن قيم المتوسطات الحسابية للعينات تقترب من قيمة المتوسط الحسابي للمجتمع الذي أخذت منه العينات. وفي مثل هذه الحالات يمكن الاستدلال على المعالم الإحصائية للمجتمع الإحصائي على نحو دقيق من خلال إحصائيات العينة.

○ التقدير الإحصائي وفترات الثقة:

يعد الاستدلال الإحصائي من الأهداف الرئيسية لعلم الإحصاء، إذ يمكن الاستعانة بالمعلومات التي توفرها العينات الجزئية المأخوذة من مجتمع إحصائي معين لتقدير المعالم الإحصائية لذلك المجتمع مثل المتوسط الحسابي والنسبة المئوية والانحراف المعياري والتباين وغيرها من المعالم الأخرى. فالتقدير الإحصائي يتمثل في اختيار قيمة أو مجموعة من القيم التي تقدر بها المعالم موضع الدراسة.

يمكن التمييز بين نوعين من التقدير اعتماداً على الهدف من وراء استخدامها، وذلك على النحو التالي:

1. **التقدير بنقطة** "Point Estimation":

ويهدف هذا النوع إلى تقدير المعلمة الإحصائية للمجتمع الإحصائي بقيمة واحدة فقط يتم حسابها من بيانات العينة. مثل تقدير المتوسط الحسابي للمجتمع من خلال المتوسط الحسابي للعينة، أو تقدير النسبة للمجتمع من خلال النسبة المتوفرة في العينة، أو تقدير الانحراف المعياري للمجتمع من خلال الانحراف المعياري للعينة. وفي هذا النوع يتم استخدام قيمة إحصائية واحدة للاستدلال على معلمة المجتمع.

2. **التقدير بفترة** "Interval Estimation":

ويهدف هذا النوع من التقدير إلى تقدير معلمة المجتمع الإحصائية بمجموعة من القيم، أي ضمن فترة من القيم تعرف باسم فترة الثقة، والتي يمكن أن تأخذ معلمة المجتمع أي قيمة ضمن حدود هذه الفترة. إن مثل هذا النوع من التقدير أكثر دقة من النوع الأول وذلك لأن التقدير بنقطة ربما لا يكون مؤشرا دقيقاً للاستدلال على معلمة المجتمع من خلال إحصائيات العينة المتوفرة. وعليه، فإن التقدير بفترة يتيح لنا إمكانية حدوث معلمة المجتمع ضمن مدى تلك الفترة.

أولاً: فترات الثقة للمتوسط الحسابي:

من أجل تقدير المتوسط الحسابي للمجتمع الإحصائي من خلال فترة من القيم حول إحصائي العينة تستخدم المعادلة التالية:

متوسط المجتمع (م) = متوسط العينة ± احتمالية الدقة المتوخاة × الخطأ المعياري

وبالرموز (م) = سَ ± ز × ع خ

أي

$$سَ \pm ز \times \frac{ع}{\sqrt{ن}}$$

حيث (ز) هي احتمال الدقة المتوخاة في التقدير، يتم تحديدها من خلال جدول التوزيع الطبيعي.

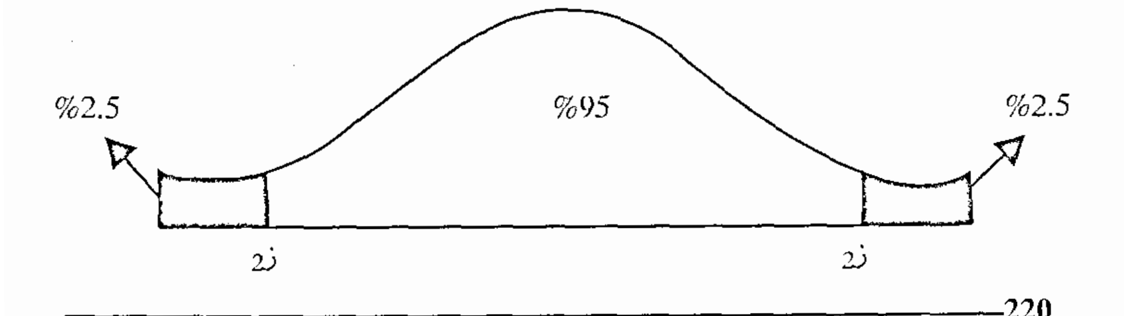
سَ : المتوسط الحسابي للعينة.

$\frac{ع}{\sqrt{ن}}$: الخطأ المعياري

مثال (8:1) عند أخذ أوزان عينة عدد عناصرها (25) شخصاً من الفئة العمرية (5 سنوات) كان المتوسط الحسابي لأوزانهم (15 كغم) بانحراف معياري لأوزان تلك الفئة = (3 كغم)، احسب فترة الثقة لتقدير المتوسط الحسابي لأوزان جميع أطفال هذه الفئة العمرية (متوسط المجتمع) بمستوى ثقة مقداره 95%.

الحل:

(1) نحدد فترة الثقة عند مستوى الثقة 95% على منحنى التوزيع الطبيعي



لاحظ أن فترة التقدير تنحصر بين قيم (ز) والتي يقع تحتها أو فوقها 2.5% من المساحة. وهذا يعني أن ($ز_2$) يقع تحتها 97.5% من المساحة. وبالنظر إلى جدول التوزيع الطبيعي نجد أن قيمة (ز) التي يقع تحتها 97.5% تساوي 1.96. وبالتعويض في المعادلة أعلاه نحصل على فترة الثقة المطلوبة.

$$(م) = \bar{س} \pm ز_{\frac{\alpha}{2}} \times \frac{ع}{\sqrt{ن}}$$

$$= 15 \pm 1.96 \times \frac{3}{\sqrt{25}}$$

$$= 15 \pm 1.96 \times \frac{3}{5}$$

$$= 15 \pm 1.96 \times 0.6$$

$$= 15 \pm 1.176$$

وعليه نلاحظ أن المتوسط الحسابي للمجتمع ينحصر بين القيمتين:

$15 - 1.176 = 13.824$

$15 + 1.176 = 15.176$ أي ضمن المدى (13.824 : 15.176).

مثال (8:2): إذا كان المتوسط الحسابي لتحصيل عينة من الطلاب عددها (16) طالباً من طلاب الصف السادس الأساسي في مادة الاجتماعيات يساوي (20) بانحراف معياري لتحصيل طلاب الصف السادس قدره (4)، فما هو المتوسط الحسابي لجميع طلاب السادس الابتدائي في مادة الاجتماعيات عند مستوى الثقة 90%.

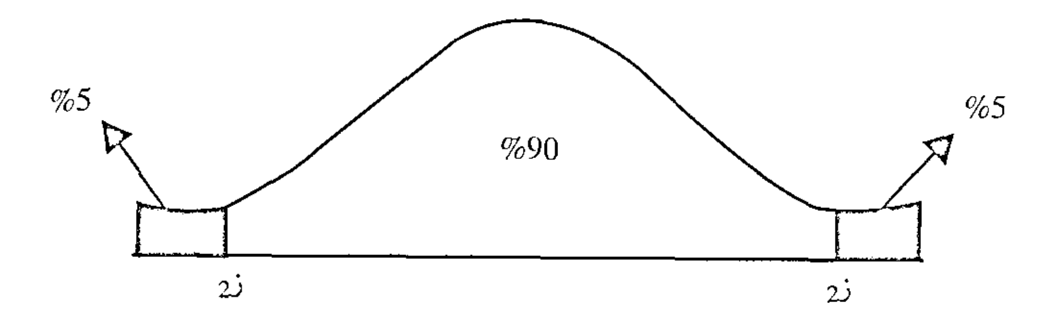
الحل:

$\alpha = 1 - 90\%$

$= 0.10$

$\frac{\alpha}{2} = 0.050$

وبالنظر إلى جدول التوزيع الطبيعي نجد أن العلامة المعيارية (ز) التي يقع تحتها 95% أو يقع فوقها 5% من المساحة تساوي (1.645).



وعليه فإن فترة الثقة للمتوسط تكون على النحو الآتي :

$$= 20 \pm 1.645 \times \frac{4}{\sqrt{16}}$$

$$= 20 \pm 1.645 \times \frac{4}{4}$$

$$= 20 \pm 1.645$$

وبالتالي نجد أن المتوسط الحسابي ينحصر بين القيمتين (18.355 & 21.645).

تجدر الإشارة هنا، أننا في حالة الأمثلة السابقة نستخدم جدول التوزيع الطبيعي وذلك لأن الانحراف المعياري للمجتمع يكون معروفاً لدينا على نحو مسبق، ولكن في أغلب الحالات مثل هذه القيمة لاتكون معروفة أو متوفرة لدينا، فعندها يستعاض عنها بحيث يستخدم بدلاً منها الانحراف المعياري لأفراد العينة، وفي مثل هذه الحالة نستخدم المعادلة التالية لتقدير الثقة :

$$\text{م} = \text{س} \pm \text{ت} \times \frac{\text{ع}}{\sqrt{\text{ن}}}$$

حيث ت : هي قيمة يتم استخراجها من جداول خاصة تعرف باسم جدول (t-students) ملحق رقم (3) وهي قيمة معادلة لقيم (ز)، وهذه القيمة تعرف باسم القيمة التائية يتطلب استخراجها تحديد مستوى الثقة المطلوب وعدد درجات الحرية (ن- 1). فعلى سبيل المثال إذا كان مستوى الثقة المطلوب هو 95% وكان عدد أفراد العينة يساوي 20، عندها ننظر في جدول التوزيع التائي تحت المساحة 97.5% والمقابلة لدرجات الحرية (ن- 1)، أي (20-1) بحيث نجد أن قيمة ت = 2.539.

مثال (8:3): ماهي قيمة (ت) إذا كان حجم العينة (25) فرداً وكان مستوى الثقة المطلوب هو 90%.

الحل: لتحديد قيمة (ت) المطلوب ننظر في جدول التوزيع التائي تحت المساحة 95% والمقابلة لدرجات الحرية (24) فنجد أن قيمة ت = 1.711 .

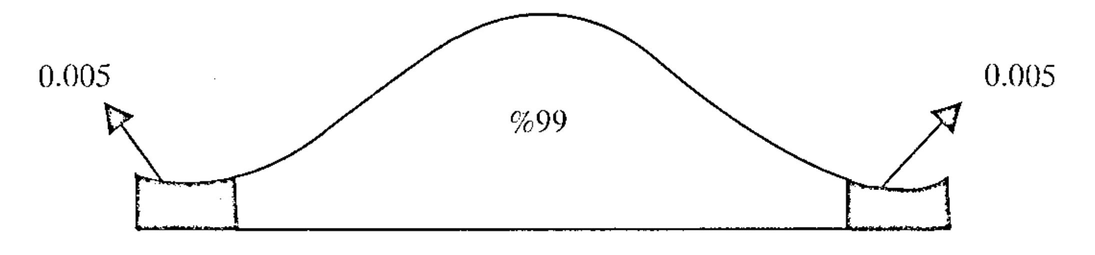
مثال (8:4): إذا علمت أن المتوسط الحسابي والانحراف المعياري لأداء أفراد عينة على مهارة معينة كان (25 ، 4) فإذا كان حجم العينة يساوي (25) فرداً، فما هي فترة الثقة لمتوسط أفراد المجتمع على تلك المهارة عند مستوى الثقة 99% .

الحل:

$$\alpha = 1 - 0.99$$

$$= 0.01$$

إذاً $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$



وعليه فإن قيمة (ت) عند درجات الحرية (25-1) ومستوى الدلالة $\frac{\alpha}{2}$ =0.005، أي المقابلة لـ 0.995 من المساحة هي (2.797).

وعليه فإن فترة الثقة تكون على النحو التالي :

$$م = 25 \pm 2.797 \times \frac{4}{\sqrt{25}}$$

$$= 25 \pm 2.797 \times \frac{4}{5}$$

$$= 25 \pm 2.797 \times 0.8$$

$$= 25 \pm 2.24$$

أي أن قيمة المتوسط الحسابي تنحصر بين القيمتين (22.76 & 27.74) .

ثانياً: فترات الثقة للتباين

يمكن تقدير التباين والانحراف المعياري لبيانات المجتمع الكلي من خلال البيانات الجزئية التي توفرها العينة، وكما هو الحال في تقدير قيم المتوسط الحسابي بنقطة أو بفترة من القيم، يمكن أيضاً تقدير التباين أو الانحراف المعياري للمجتمع بقيمة واحدة مثل اعتبار الانحراف المعياري مساوياً لقيمة الانحراف المعياري للعينة (σ =ع) أو اعتبار التباين مساوياً للتباين للعينة (σ^2 =$ع^2$). ولكن هذا التقدير قد لايكون دقيقاً، إذ ليس من الضروري أن يكون تباين العينة المسحوبة مساوياً لتباين المجتمع.

وعوضاً عن ذلك ولضمان توخي الدقة يمكن تقدير التباين للمجتمع من خلال فترة من القيم تعرف بفترات الثقة للتباين.

يتطلب تحديد فترات الثقة للتباين حساب التباين لبيانات العينة وتحديد قيمة (كاي^2 : كاي تربيع ويرمز لها بالرمز χ^2) عند مستوى الثقة المطلوب ودرجات الحرية التي هي بمثابة (حجم العينة – 1). يمثل كاي^2 (مربع كاي) التوزيع العيني للنسب بين تباين العينات التي يمكن أخذها من مجتمع ما وتباين ذلك المجتمع، ومثل هذه القيمة يمكن إيجادها من خلال جدول خاص يعرف بجدول توزيع كاي^2 ملحق رقم (4). عموماً يمكن استخدام المعادلة التالية لتحديد فترات الثقة لتباين المجتمع:

$$\frac{\acute{ع}^2}{\text{كاي}^2_{\frac{\alpha}{2}} / \text{د.ح}} \leq ع^2 \leq \frac{\acute{ع}^2}{\text{كاي}^2_{\frac{\alpha-1}{2}} / \text{د.ح}}$$

كاي$^2_{\frac{\alpha}{2}}$ / د.ح : هي عبارة عن ناتج قسمة قيمة كاي^2 على حجم العينة.

مثال (8:5): ماهي قيمة كاي^2 عند مستوى الدلالة = 0.05 ودرجات حرية (30).

الحل: α = 0.05

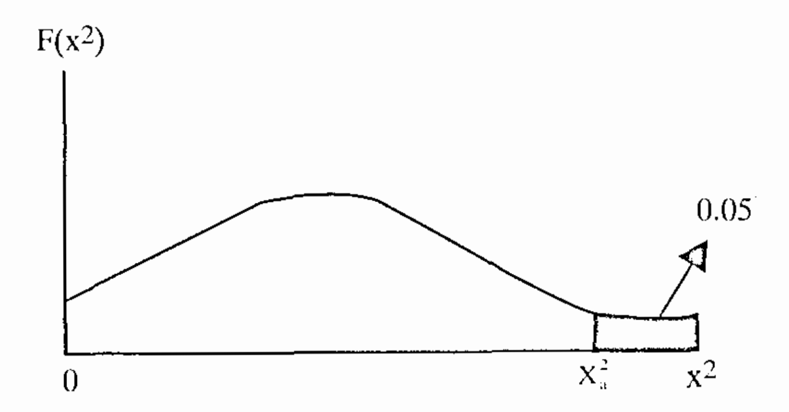
درجات الحرية = (ن – 1)

= (1-30)

= 29

إذاً كاي^2: (α = 0.05 , 29) من جدول توزيع كاي^2: ملحق رقم (4)

= 42.5569



مثال (8:6): ماهي قيمة كاي^2 عند مستوى الثقة 95% ودرجات حرية (30).

الحل:

د.ح= 29 مستوى الثقة 95%.

بالنظر إلى جدول توزيع كاي^2: (x^2)

تحت مستوى الثقة 95% والمقابل لدرجات الحرية 29 نجد أن قيمة:

كاي^2: (x^2) = 17.7083

مثال (8:7): عند أخذ عينة من الطلاب عدد أفرادها 25 طالباً وجد أن التباين لتحصيل هؤلاء الطلاب في مادة الرياضيات = (20)، احسب فترات الثقة لتباين جميع الطلاب في مادة الرياضيات عند مستوى الثقة 95%.

الحل:

أولاً: نجد قيمة كاي^2 (x^2) عند مستوى الدلالة (1- مستوى الثقة): $\frac{\alpha}{2}$ ودرجات الحرية (ن-1).

مستوى الدلالة (α) = 1- مستوى الثقة

= 1 - 0.95

= 0.05

$\frac{\alpha}{2} = 0.025$ ، د.ح = (25-1)

من الجدول (كاي^2 عند $\frac{\alpha}{2}$ =0.025 ، د.ح=24) = 39.3641.

ثانياً: نجد قيمة كاي^2: (X^2) عند مستوى ($1 - \frac{\alpha}{2}$)

أي عند مستوى الدلالة $1 - \frac{0.05}{2}$ ودرجات حرية قدرها (24)،

أي عند مستوى الثقة 0.975 ودرجات حرية (24).

إذاً قيمة كاي^2 : (X^2) عند درجات الحرية (25-1) ومستوى الثقة 0.975 هي 12.4011 ، وعليه فإن فترات الثقة للتباين تكون على النحو الآتي:

$$\frac{20}{25 \div 39.3641} \leq ع^2 \leq \frac{20}{25 \div 12.4011}$$

$$\frac{20}{1.575} \leq ع^2 \leq \frac{20}{0.4960}$$

$$12.70 \leq ع^2 \leq 40.32$$

وهكذا نلاحظ أن قيمة التباين للمجتمع الكلي تنحصر بين القيمتين (12.70 & 40.32).

مثال (8:8) إذا كان التباين لأداء أفراد عينة عدد أفرادها (ن = 20) يساوي 15 ، احسب فترات الثقة لأداء مجموعة الأفراد التي تنتمي إلى هذه العينة عند مستوى الثقة 99%.

الحل:

$$\frac{\alpha}{2} = \frac{1-0.099}{2} = \frac{0.01}{2} = 0.005$$

$$1 - \frac{\alpha}{2} = 1 - \frac{0.01}{2}$$

$$= 1 - 0.005 = 0.995$$

إذا قيمة كاي^2 : (X^2) عند $\frac{\alpha}{2}$ = 0.005 و درجات حرية = 19 هي (38.5822) .

أما قيمة كاي^2: (X^2) عند $1-\frac{\alpha}{2}$ =0.995 ودرجات حرية = 19 فهي (6.84398) .

وعليه فإن فترة الثقة للتباين المطلوب تكون :

$$\frac{15}{20 \div 38.5822} \leq ع^2 \leq \frac{15}{20 \div 6.84398}$$

$$\frac{15}{1.92911} \leq ع^2 \leq \frac{15}{0.3422}$$

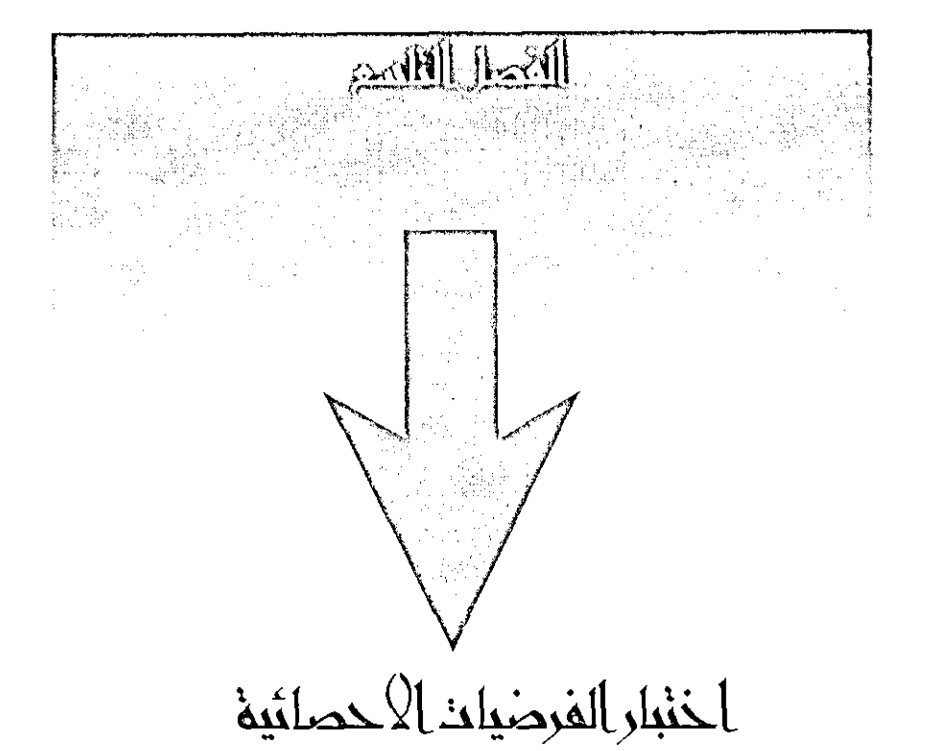
$$7.776 \leq ع^2 \leq 43.83$$

وهكذا نجد أن التباين ينحصر بين القيمتين

[43.83-7.776]

○ كلمة أخيرة في هذا الموضوع

إن تقدير معالم المجتمع من خلال إحصائيات العينة سواءاً بقيمة واحدة، أو بفترة من القيم (فترات الثقة) لا يقتصر فقط على المتوسط الحسابي والتباين، وإنما يتعدى ذلك ليشمل تقدير معالم أخرى للمجتمع الإحصائي مثل معامل الارتباط والنسبة، وكذلك يتضمن تقدير الفرق بين متوسطين حسابيين لمجتمعين إحصائيين والفرق بين تباين مجتمعين على ظاهرة معينة وكذلك الفرق بين النسب ومعاملات الارتباط. وكون أن هذه المواضيع خارج نطاق هذا الكتاب فلم يتم التعرض لها.



اختبار الفرضيات الاحصائية

Statistical Hypothesis Testing

تمهيد:

لقد مر معنا سابقاً أن من اهداف علم الإحصاء هو الاستدلال حول معالم المجتمع الاحصائي من خلال البيانات الجزئية التي توفرها لنا العينات المسحوبة من ذلك المجتمع . وقد عرفنا في الفصل السابق ان فترات الثقة هي من إحدى الاجراءات الاحصائية المستخدمة في تقدير معالم المجتمع مثل المتوسط الحسابي والانحراف المعياري وغيرها من المعالم الأخرى في ضوء مستوى ثقة محدد يعرف بمستوى دقة التقدير . وفي هذا الفصل سيتم التعرض الى أجراء احصائي آخر يعرف باختبار الفرضيات يمكن من خلاله تقدير معلمه احصائية أو أكثر من معالم المجتمع الاحصائية .

ففي الكثير من الحالات لا تتوفر لدينا معلومات كافية عن مجتمع احصائي معين او ظاهرة ما ، فعندها يتم اللجوء إلى الاسلوب العيني ، بحيث يتم دراسة مجموعة جزئية من ذلك المجتمع بهدف اصدار أحكام او اتخاذ قرارات حول المجتمع ككل . ولضمان دقة التعميم بالنتائج من خلال البيانات الجزئية (العينة) على البيانات الكلية (المجتمع) لا بد من اختبار مدى دقة الاحصائيات التي توفرها بيانات العينة في تقدير المعالم الاحصائية للمجتمع الذي اخذت منه العينة . وفي مثل هذه الحالة يمكن اللجوء الى استخدام اختبار الفرضيات . فعلى سبيل المثال إذاكان لدينا مجتمع من الافراد يبلغ تعداده (10000) فرد وكان الهدف معرفة المتوسط الحسابي لاوزان جميع الافراد في ذلك المجتمع ، فإذا علمنا أن هناك عوائق تمنع اخذ اوزان جميع هؤلاء الافراد ، فعندها يمكن اخذ مجموعة جزئية بطريقة عشوائية من ذلك المجتمع تسمى بالعينة ، بحيث يصار الى قياس اوزان افرادها وتحديد المتوسط الحسابي والانحراف المعياري لاوزانهم ، ومثل هذه الاحصائيات تستخدم للتنبؤ او تقدير الاحصائيات المناظرة لها في المجتمع ككل . اذ يتم اختبار دقتها في التنبؤ أو التقدير من خلال فحص الفرضية الاحصائية .

تعرف الفرضية الاحصائية على أنها صياغة مبدئية حول معلمه أو أكثر من معالم المجتمع الاحصائي ، ويمكن ان تكون صحيحة أو خاطئة بحيث نلجأ إليها لتقدير معالم المجتمع الاحصائي عندما لا تتوفر لدينا معلومات كافية عن المجتمع الاحصائي . فالفرضية الاحصائية تعد بمثابة افضل تخمين او تنبؤ لمعالم المجتمع من خلال احصائيات العينة ، وهي بالتالي تساعد الباحثين على اصدار بعض الاحكام او اتخاذ القرارات المناسبة حول المعالم الاحصائية في ظل البيانات المتوفرة .

يستخدم اجراء اختبار الفرضيات في معظم البحوث التجريبية وشبه التجريبية ، اذ يلجأ الباحثون على نحو مسبق على صياغة افتراض معين حول ظاهرة ما ويعملون على جمع البيانات الاحصائية المناسبة من اجل الحكم على مدى صحة هذا الافتراض او عدم صحته ، وعادة يسمى الافتراض الذي يصيغه الباحث حول الظاهرة بالفرضية التجريبية Experimental Hypothesis وهي تمثل النواتج المتوقعة للبحث اوالدراسة . ومثل هذه الفرضية بالتالي قد تدعم صحتها البيانات المتوفرة او تبرهن عدم صحتها في ضوء ذلك بعد اجراء عمليات المقارنة والتقييم لهذه الفرضية (Christensen & Stoup, 1986) .

عند اللجوء الى استخدام اجراء اختبار الفرضيات الاحصائية ، يتطلب الامر دائما صياغة نوعين من الفرضيات هي :

أولاً : الفرضية الصفرية :Null Hypothesis : وتعرف باسم الفرضية الصفرية أو العدمية ويرمز لها بالرمز (H0)، ومثل هذه الفرضية تنفي وجود اية فروق بين احصائيات العينة واحصائيات المجتمع ، او انها تنفي وجود اثر لمتغير في متغير اخر او علاقة بين ظاهرتين . وتكتب احصائياً على النحو التالي :

$$H0: X = \mu$$

حيث X : المتوسط الحسابي للعينة

μ : المتوسط الحسابي للمجتمع

ولفظياً يمكن التعبير عن هذه الفرضية على النحو التالي :

لا يوجد فرق ذو دلالة احصائية عند مستوى دلالة معينة (α=0.05 أو 0.01 مثلاً) بين المتوسط الحسابي للعينة والمتوسط الحسابي للمجتمع .

وفي حالة دراسة اثر متغير في متغير اخر ، يمكن التعبير عن الفرضية الصفرية على النحو التالي :

H_0 : لا يوجد فروق ذو دلالة احصائية بين متوسط اداء الذكور على المتغير (س) والمتوسط الحسابي لاداء الإناث على ذلك المتغير.

$$H0: \mu_1 = \mu_2$$

حيث : μ_1 : المتوسط الحسابي لاداء الذكور

μ_2 : المتوسط الحسابي لاداء الاناث.

مثال (9:1): لنفرض ان باحثاً اراد أن يختبر اثر التعزيز في تحصيل طلبة فصل دراسي معين في مادة الرياضيات، فعندها يمكن أن تكتب الفرضية الصفرية على النحو التالي:

$$H_0: \mu_1 = \mu_2$$

وهذا يعني ليس للتعزيز اثراً في التحصيل في مادة الرياضيات؛ أي ان تحصيل الافراد الذين تلقوا التعزيز مساوٍ لتحصيل الافراد الذين لم يتلقوا التعزيز في مادة الرياضيات.

مثال (9:2): أراد باحث ان يقارن بين اداء الاناث والذكور في ادراك المعنى اللغوي للمفردات الغامضة. وعليه فإن الفرضية الصفرية (H_0) يمكن كتابتها على النحو التالي

$$H_0: \mu_1 = \mu_2$$

أي انه لا يوجد فرق ذو دلالة احصائية بين متوسط اداء الذكور في إدراك المعنى اللغوي للمفرادت الغامضة ومتوسط اداء الاناث في إدراك المعنى. وبلغة اخرى لا يختلف ادراك الذكور عن ادراك الاناث لمعاني المفردات الغامضة.

ثانياً: الفرضية البديلة Alternative Hypothesis: وهي الفرضية الاحصائية التي تتعارض مع الفرضية الصفرية والتي يسعى الباحث الى اثبات صحتها من خلال البيانات المتوفرة لديه. ومثل هذه الفرضية قد تكون غير متجهة (ذات طرفين) او متجهة (ذات طرف واحد). ويرمز لمثل هذه الفرضية بالرمز (H_1).

فعندما تكون الفرضية البديلة غير متجهة فهي تكتب على النحو التالي:

$$H_1: \mu_1 \neq \mu_2$$

هذا يعني ان المتوسط الحسابي الاول لا يساوي المتوسط الحسابي الثاني؛ أي ان هناك فرقاً ذو دلالة احصائية بين المتوسط الحسابي الاول والمتوسط الحسابي الثاني:

أما عندما تكون الفرضية البديلة متجهة ، فهي تكتب على النحو التالي :

$$(1) ....H_1: \mu_1 > \mu_2$$

$$(2) ....H_1: \mu_1 < \mu_2$$

ففي الحالة الاولى يعني ان المتوسط الحسابي الاول اكبر من المتوسط الحسابي الثاني ، في حين في الحالة الثانية يعني ان المتوسط الحسابي الاول اصغر من المتوسط الحسابي الثاني :

مثال(9:3) : في دراسة ما أراد باحث ان يختبر دلالة الفرق بين المتوسط الحسابي لمعامل ذكاء كل من الذكور والاناث في مجتمع معين ، فما هي الفرضية الصفرية والفرضية البديلة لهذه الدراسة .

الحل :

أ. الفرضية الصفرية H_0 :

$$H_0: \mu_1 = \mu_2$$

لفظياً : لا يوجد فروق ذات دلالة احصائية بين المتوسط الحسابي لمعامل ذكاء الذكور والمتوسط الحسابي لمعامل ذكاء الاناث ؛ أي انهما متساويان .

ب- الفرضية البديلة H1 :

1- اذا كانت متجهة

$$H_1: \mu_1 > \mu_2$$

لفظياً : المتوسط الحسابي لمعامل ذكاء الذكور اكبر من المتوسط الحساب لمعامل ذكاء الاناث .

أو

$$H_1: \mu_1 < \mu_2$$

لفظياً : المتوسط الحسابي لمعامل ذكاء الذكور اصغر من المتوسط الحسابي لمعامل ذكاء الاناث.

2- اذا كانت غير متجهة

$$H_1: \mu_1 \neq \mu_2$$

لفظياً: المتوسط الحسابي لمعامل ذكاء الذكور لا يساوي المتوسط الحسابي لمتوسط معامل ذكاء الاناث؟

*** مستوى الدلالة Level of Significance**

عند اللجوء إلى استخدام اجراء اختبار الفرضيات يتطلب الامر تحديد محك معين لتقدير الدقة في عمليات التقدير او المقارنة. ومثل هذا المحك يعرف بمستوى الدلالة للاختبار.

ففي عملية الاختبار قد نقع احياناً في نوعية من الاخطاء، بحيث نرتكب الخطأ الاول عندما نقرر رفض الفرضية الصفرية وهي في الواقع صحيحة، ويعرف هذا النوع من الخطأ بالخطأ من النوع الاول ويرمز له بالرمز ،(α)، أما الخطأ الثاني فنقع فيه عندما نقرر قبول الفرضية الصفرية وهي في حقيقة الامر خاطئة، ويعرف هذا النوع من الخطأ بالخطأ من النوع الثاني، ويرمز له بالرمز (B).

*** خطوات اختبار الفرضيات**

تتطلب عملية اختبار الفرضيات الخطوات التالية:

1- تحديد وصياغة كل من الفرضية الصفرية والفرضية البديلة.

يعتمد صياغة الفرضية البديلة على طبيعة البحث والاهداف التي يسعى الباحث الى تحقيقها، فقد تكتب بصيغة غير متجهة أو بصيغة متجهة كما أسلفنا سابقاً.

2- تحديد مستوى الدلالة المطلوب

يعد مستوى الدلالة المحك المعنوي الذي يمثل احتمال رفض الفرضية الصفرية وهي في الواقع صحيحة ويرمز له بالرمز (α). يجب أن يكون حجم(α) صغيراً جداً، إذا ان معظم الاحصائيين يؤكدون ضرورة أن لا يتجاوز حجم (α) الـ(0.05 أو 01.) وذلك لان الهدف هو اثبات صحة الفرضية البديلة، لذلك كلما كان احتمال رفض الفرضية الصفرية وهي صحيحة منخفضاً، كان التقدير الاحصائي اكثر دقة. فعلى سبيل المثال، إذا كان التقدير المطلوب عند مستوى الدلالة (α=0.01)، فهذا يعني ان احتمال رفض الفرضية الصفرية بالرغم أنها صحيحة يساوي1%؛ أي ان احتمال اثبات صحتها يساوي 99%.

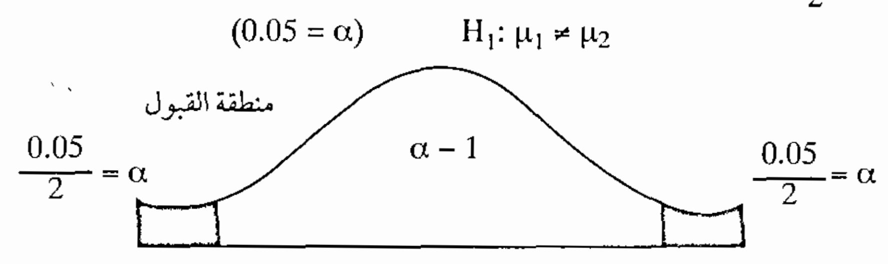
3- تحديد نوع الاختبار الاحصائي المطلوب

هناك انواعاً عديدة من الاختبارات الاحصائية مثل اختبار (Z) واختبار(t) (كاي^2) واختبار (ف) وغيرها من الاختبارات الاحصائية الأخرى. ويتحدد نوع الاختبار الاحصائي المستخدم في ضوء طبيعة

التوزيع العيني والهدف الاحصائي بالاضافة الى مدى توفر الافتراضات الرئيسية التي تتبع استخدام الاحصائي المناسب. وفي هذا الفصل سوف يتم التعرض فقط لكل من اختبار (Z) واختبار (t) فقط.

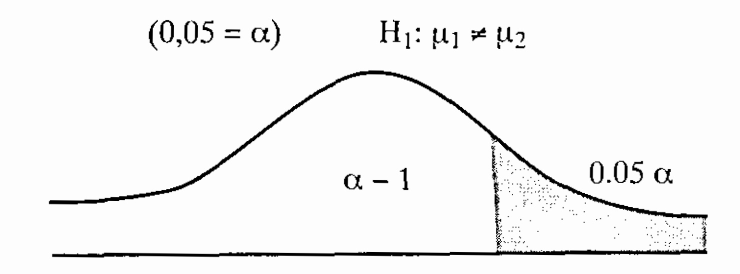
4 -تحديد منطقة الرفض والقبول

يتم عادة تحديد منطقة الرفض والقبول اعتماداً على مستوى الدلالة المطلوبة واتجاه الفرضية البديلة. ففي حالة كون الفرضية البديلة غير متجهة $(H_1:\mu_1=\mu_2)$ تكون منطقة الرفض على طرفي التوزيع والتي تقابل $\frac{\alpha}{2}$ ، وذلك كما هو موضح بالشكل التالي:

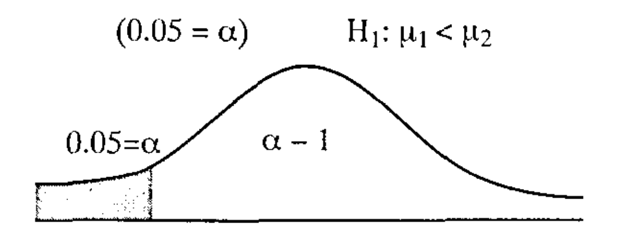


في الشكل اعلاه: نلاحظ أن منطقة الرفض وقعت على ذيلي التوزيع (المنطقة المضللة)، وهي عبارة عن المساحة التي تقع فوق أو تحت العلامة المعيارية (ز) المقابلة $\frac{\alpha}{2}$.

أما عندما تكون الفرضية البديلة متجهة، فأن منطقة الرفض تكون على احد ذيلي التوزيع، وهي تمثل المساحة التي تقع فوق او تحت مستوى الدلالة (α) . حسب صيغه الفرضية البديله. كما هو موضح بالشكل التالي:



في الشكل اعلاه نلاحظ أن منطقة الرفض تمثل المساحة التي تقع فوق العلامة المعيارية (ز) والمقابلة لـ (0.05)=(α)



في الشكل اعلاه نلاحظ أن منطقة الرفض تمثل المساحة التي تقع تحت الدرجة المعيارية (−ز) والمقابلة لمستوى الدلالة (α =0.05).

5- حساب قيمة الاحصائى باستخدام البيانات المتوفرة. يتم عادة حساب قيمة الاحصائي من خلال معادلة احصائية خاصة اعتماداً على طبيعة البيانات المتوفرة والهدف من اختبار الفرضيات. هذا وسيصار الى التعرف الى بعض هذه المعادلات لاحقاً.

6- مقارنة قيمة الاحصائى المحسوبة مع قيمة الاحصائي المرتبط بمنطقة الرفض والذي على اساسه يتم اتخاذ قرار حول رفض او قبول الفرضية الصفرية.

*** اختبارات (Z) و (T)**

تستخدم هذه الاختبارات لاختبار الفرضيات وهي تعطي نفس النتائج تقريباً. ويتوقف الاختلاف فيما بينها من حيث الاستخدام فقط. فالاختبار الزائي (Z) يستخدم في حالة توفر بعض المعلومات عن معالم المجتمع الاحصائى مثل الانحراف المعياري، في حين يتم اللجوء الى الاختبار (T) عندما لا يكون لدينا اية معلومات عن الانحراف المعياري للمجتمع. وفيما يلي عرض تفصيلي مع الامثلة عن هذه الاختبارات.

أولاً: اختبار المتوسط الحسابي في حالة معرفة الانحراف المعياري: يستخدم اختبار (Z) لاختبار المتوسط الحسابي لمجتمع احصائي معين من خلال البيانات التي توفرها العينة وذلك في حالة توفر معرفة مسبقة حول الانحراف المعياري للمجتمع الاصلي الذي سحبت منه العينة، ويتم ذلك من خلال المعادلة التالي:

$$\text{ز} = \frac{\bar{\text{س}} - \bar{\text{س}}_{\text{م}}}{\text{ع} / \sqrt{\text{ن}}}$$

حيث سَ : المتوسط الحسابي للعينة

س م : المتوسط الحسابي للمجتمع

ع : الانحراف المعياري للمجتمع

ن : عدد افراد العينة

مثال(9:4) : ينتج مصنع قضباناً حديدية بأطوال مختلفة ، قام باحث بأخذ عينة من القضبان عددها (64) قضيباً وحسب المتوسط الحسابي لاطوالها فكان (5.5) متر ، فإذا كان الانحراف المعياري لاطوال القضبان التي ينتجها المصنع يساوي (2) متر .

عند مستوى الدلالة $\alpha = 0.05$ جد ما يلي :

أ- اختبر الفرضية الصفرية التي تنص على أن المتوسط الحسابي لاطوال القضبان يساوي (5) متر مقابل الفرضية البديلة التي تنص على أن المتوسط الحسابي لاطوال القضبان لا يساوي(5) متر .

ب-اختبرالفرضية الصفرية التي تنص على ان المتوسط الحسابي لاطوال القضبان يساوي (5) متر مقابل الفرضية البديلة التي تنص على ان المتوسط الحسابي لاطوال القضبان اكبر من (5) متر .

جـ- اختبر الفرضية الصفرية التي تنص على أن المتوسط الحسابي لاطوال القضبان يساوي (5) متر مقابل الفرضية البديلة التي تنص على أن المتوسط الحسابي لاطوال القضبان اصغر من (5) متر .

الحل :

أولاً: إجابة الفرع (أ) من السؤال:

$$H_0: \mu = 5$$

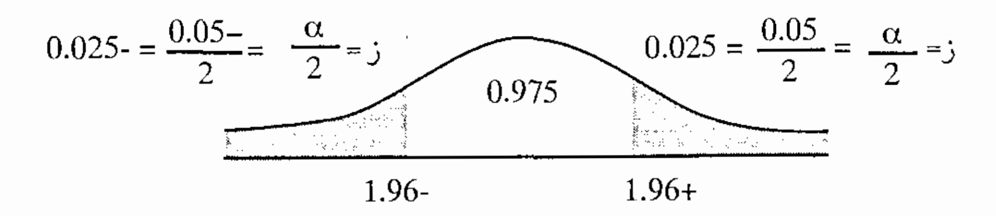
$$H_1: \mu \neq 5$$

$$ز = \frac{سَ - سَ_م}{ع / \sqrt{ن}} = \frac{5-5.5}{2 / \sqrt{64}}$$

$$= \frac{0.5+}{0.25} = 2+$$

إذا قيمة (ز) المحسوبة تساوي +2

منطقة الرفض (قيمة ز الحرجة) هي :



من جدول التوزيع الطبيعي نجد ان قيمة ز= 1.96

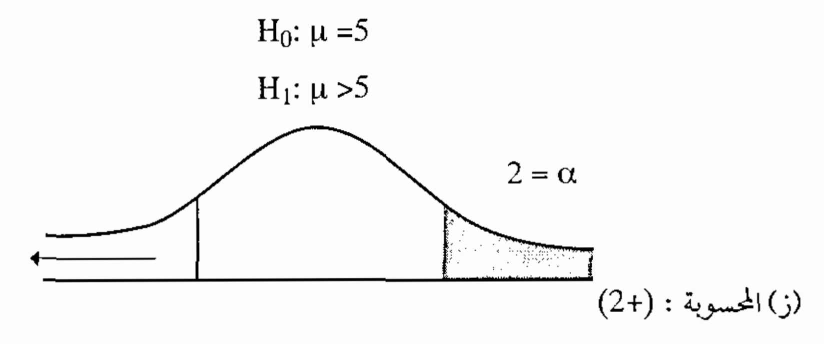
وعليه فأننا نرفض الفرضية الصفرية اذا كانت قيمة (ز) المحسوبة اكبر من قيمة ز= +1.96 أو اقل من قيمة ز= -1.96

بما أن قيمة (ز) المحسوبة = (+2) وهي أكبر من قيمة (ز) الحرجة (+1.96) ، فالقرار هو رفض الفرضية الصفرية ، اي ان البيانات المتوفرة غير كافية لدعم الفرضية الصفرية التي تنص على أن المتوسط الحسابي لاطوال القضبان التي ينتجها المصنع يساوي (5) متر.

ثانياً: اجابة الفرع (ب) من السؤال.

$H_0: \mu = 5$

$H_1: \mu > 5$



من جدول التوزيع الطبيعي نجد أن قيمة (ز) التي يقع فوقها (0.5) من المساحة تساوي (1.645). ففي حالة مثالنا هذا فأننا نرفض الفرضية الصفرية إذا كانت قيمة (ز) المحسوبة أكبر من قيمة (ز) الحرجة التي تساوي 1.645 . واعتماداً على ذلك فإن القرار هو رفض الفرضية الصفرية، إذ يمكن القول ان البيانات المتوفرة لا تدعم الفرض الذي ينص على ان متوسط اطوال القضبان التي ينتجها المصنع يساوي (5) متر . وهذا يعني أن متوسط آطوال القضبان هو أطول من (5) متر.

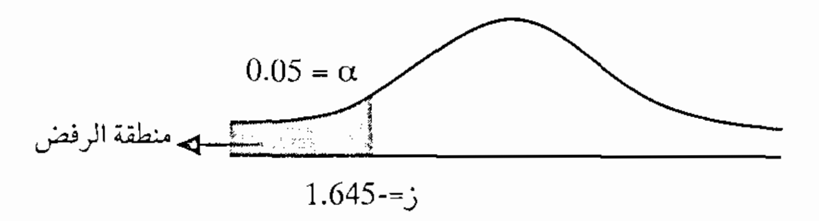
ثالثاً: اجابة الفرع (جـ) من الســؤال

$$H_0: \mu=5$$

$$H_1: \mu<5$$

قيمة (ز) المحسوبة = (2+)

قيمة (ز) الحرجة ←منطقة الرفض



في هذه الحالة نرفض الفرضية الصفرية اذا كانت قيمة (ز) المحسوبة اقل من قيمة (ز) الحرجة. ولما كانت قيمة (ز) المحسوبة تساوي (2+) وهي أكبر من قيمة (ز) الحرجة التي تساوي (-1.645)، فعليه فأن القرار هو قبول الفرضية الصفرية، اي ان البيانات المتوفرة غير كافية لرفض الفرضية الصفرية، بمعنى أن المتوسط الحسابي لاطوال القضبان التي ينتجها المصنع تساوي أو آطول من (5) متر.

ثانياً: اختبار المتوسط الحسابي في حالة عدم معرفة الانحراف المعياري:

عندما لا تتوفر لدينا اية معلومات عن الانحراف المعياري للمجتمع الاحصائي، عندها يتم الاستعاضة عن الاختبار الاحصائي (Z) ليستخدم عوضاً عنه الاختبار الاحصائي(T). وفي مثل هذه الحالة يتم استخدام الانحراف المعياري التي يتم حسابه من بيانات العينه، بحيث يتم حساب قيمة الاحصائي(T). وفي مثل هذه الحالة يتم استخدام الانحراف المعياري التي يتم حسابه من بيانات العينه، بحيث يتم حساب قيمة الاحصائي (T) وفق المعادلة التالية:

$$ت = \frac{سَ - سَ_م}{ع / \sqrt{ن}}$$

حيث: سَ: المتوسط الحسابي للعينة

سَ م: المتوسط الحسابي للمجتمع

عَ: الانحراف المعياري للعينة

ن : عدد افراد العينة

وهنا يتحدد قبول الفرضية الصفرية او عدم قبولها في ضوء مقارنة قيمة (T) المحسوبة مع قيمة (T) الحرجة عند مستوى دلالة معين ودرجات حرية تتمثل في (ن-1).

مثال (9:5) قام باحث بأخذ عينة من الاطفال ممن تتراوح اعمارهم بين (7-9) سنوات وكان عدد افراد هذه العينة (25) طفلاً فوجد أن المتوسط الحسابي لاطوالهم يساوي (120سم) بانحراف معياري مقداره (15سم). عند مستوى الدلالة (α =0.01) اجب عن الاسئلة التالية:

أ- اختبر الفرضية الصفرية $H_0: \mu=100$ مقابل الفرضية البديلة $H_1:\mu \neq 100$.

ب- اختبر الفرضية الصفرية $H_0:\mu=100$ مقابل الفرضية البديلة $H_1: \mu>100$

جـ- اختبر الفرضية الصفرية $H_0:\mu=100$ مقابل الفرضية البديلة $H_1:\mu<100$

الحل:

1- اجابة الفرع (أ) من السؤال : $H_1:\mu_1= 100$ مقابل $H_1: \mu_1 \neq 100$

ت المحسوبة= $\frac{100 - 120}{\sqrt{25}/15} = \frac{20}{5/15} = \frac{20}{3} \approx 6.67$

منطقة الرفض = $\frac{\alpha}{2} = \frac{0.01}{2}$

قيمة (ت) الحرجة ∈ منطقة الرفض

من جدول توزيع (T) نجد فيه (T) الحرجة عند مستوى الدلالة $\frac{\alpha}{2} = \frac{0.1}{2}$

ودرجات الحرية (ن-1 : 25-1) وهي تساوي 2,797.

في حالة هذه الفرضية نرفض الفرضية الصفرية اذا كان مطلق قيمة (ت) المحسوبة |ت| اكبر من قيمة (ت) الحرجة. وبالنظر الى قيمة |ت| المحسوبة والحرجة. نجد أن قيمة ت المحسوبة اكبر من قيمة (ت) الحرجة، وعليه فإن القرار هو رفض الفرضية الصفرية، إذ إن البيانات المتوفرة غير كافية لدعم صحة الفرضية الصفرية. وهذا يعنى أن متوسط أطوال الأطفال قد يكون أطول أو أقصر من 100سم.

2- اجابة الفرع (ب) من السؤال : $H_0{:}\mu=100$ مقابل $H_1{:}\mu>100$

قيمة ت المحسوبة = 6.67

قيمة ت الحرجة عند مستوى الدلالة ($\alpha = 0.01$) ودرجات حرية (24) هي 2.492 . في حالة مثالنا هذا نرفض الفرضية الصفرية اذا كانت قيمة (ت) المحسوبة اكبر من قيمة (ت) الحرجة . وبما ان قيمة (ت) المحسوبة (6.67) أكبر من قيمة (ت) الحرجة (2.492)، إذاً فالقرار هو رفض الفرضية الصفرية لصالح الفرضية البديلة . وهذا يعني أن متوسط أطوال الأطفال هو أكبر من (100) سم .

3- اجابة الفرع (جـ) من السؤال : $H_0{:}\ \mu=100$ مقابل $H_1{:}\mu<100$.

قيمة (ت) المحسوبة = 6.67

قيمة (ت) الحرجة = 2.492

في هذه الحالة نرفض الفرضية الصفرية اذا كانت قيمة (ت) المحسوبة اقل من (– ت) الحرجة . وبما أن قيمة (ت) المحسوبة أكبر من قيمة (–ت) الحرجة فالقرار هو قبول الفرضية الصفرية، اي ان البيانات المتوفرة غير كافية لرفض الفرضية الصفرية . وهذا يعني أن أطوال الأطفال يساوي أو أكبر من (100)سم .

ثالثاً: اختبار النسبه لمجتمع احصائي

كما هو الحال في اختبار الفرضيات المتعلقة بالمتوسط الحسابي، يمكن ايضاً اختبار الفرضيات التي تتعلق بنسبة معينة لمجتمع احصائي معين، ويتم حساب قيمة الاحصائي (Z) في مثل هذه الحالة من خلال المعادلة التالية :

$$Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0\,(1 - P_0)}{ن}}}$$

حيث $\hat{p}$: هي النسبة داخل العينة وهي تساوي عدد الحالات في العينة مقسوماً على عدد افراد العينة $\left(\frac{X}{ن}\right)$

P_0 : وهي عبارة عن ناتج $(1 - \hat{P})$

ن: عدد الافراد في العينة

مثال (9:6) : قام باحث بأخذ عينه بلغ عدد افرادها (900) شخص وكان يهدف الى معرفة وجهة نظر هؤلاء الافراد حول عمل المرأة، فوجد ان (600) فرداً منهم يؤيدون عمل المرأة. اختبر ما اذا كانت نسبه المؤيدين تختلف عن70% عند مستوى الدلالة α=0.05 ؟

الحل: $H_0: P= 0.70$ مقابل $H_1: P \neq 0.70$

$$\widehat{P} = \frac{X}{ن} = \frac{600}{900} = 0.67$$

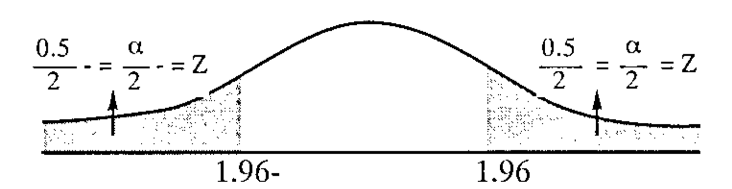
$$P_0 = 1-\widehat{P} = 1- 0.67 = 0.33$$

$$Z = \frac{0.70 - 0.67}{\sqrt{\frac{0.33\ (1 - 0.33)}{900}}}$$

$$Z = \frac{0.03}{\sqrt{\frac{0.2211}{900}}} = \frac{0.03}{\sqrt{0.00025}} = \frac{0.03}{0.0158} \approx 1.90$$

إذا قيمة (z) المحسوبة= 1.90

قيمة (Z) الحرجة عند مستوى الدلالة $\frac{\alpha}{2}$ هي 1.96



القرار: نرفض الفرضية الصفرية اذا كان مطلق قيمة (Z) المحسوبة $|Z|$ * أكبر من قيمة (Z) الحرجة. وبما ان مطلق قيمة (Z) المحسوبة=1.90 وهو اقل من قيمة (Z) الحرجة، فالقرار هو قبول الفرضية الصفرية، أي ان البيانات المتوفرة تؤكد عدم اختلاف النسبه في العينه عن 70% .

مثال (9:7) : رغب معلم في دراسة نسبة النجاح لدى المتقدمين لامتحان الشهادة الثانوية في محافظة اربد، قام بأخذ عينه من المتقدمين لذلك الامتحان بلغ عدد افرادها (1200) طالباً ووجد أن 720 طالباً منهم قد اجتازوا الثانوية العامة بنجاح. اختبر اذا ما كانت نسبة النجاح تتجاوز نسبة 65% عند مستوى الدلالة α=0.01 ؟

* $|Z|$ مطلق قيمة (Z).

الحل:

$$H_0: P = 0.65$$

$$H_1: P > 0.65$$

$$\hat{P} = \frac{X}{?} = \frac{700}{1200} = 0.60$$

$$P = 1 - \hat{P} = 1 - 0.60 = 0.40$$

$$Z = \frac{0.06 - 0.65}{\sqrt{\frac{0.40\,(1 - 0.40)}{1200}}}$$

$$Z = \frac{-0.05}{\sqrt{\frac{0.24}{1200}}} = \frac{-0.05}{\sqrt{0.0002}} = \frac{-0.05}{0.0141} \approx -3.55$$

منطقة الرفض

$0.01 = \alpha$

$Z = 2.33$

القرار: نرفض الفرضية الصفرية اذا كانت قيمة (Z) المحسوبة اكبر من قيمة (Z) الحرجة. وبما ان قيمة (Z) المحسوبة اقل من قيمة (Z) الحرجة، فالقرار هو عدم رفض الفرضية الصفرية، اي ان البيانات المتوفرة تؤكد أن النسبة لا تتجاوز نسبة 65% .

رابعاً: اختبار الفرق بين متوسطين لعينتين مستقلتين: في بعض المواقف يتم اخذ عينتين مستقلتين عن بعضهما البعض من مجتمع احصائي واحد أو من مجتمعين احصائيين مستقلين، بهدف دراسة اثر متغير او اكثر في ظاهرة معينة لدى هاتين العينتين او المقارنة بينهما وفقاً لابعاد معينة، او الاستدلال حول متوسطي المجتمعين من خلال متوسطي العينتين . ففي مثل هذه الحالات وغيرها يتم اللجوء عادة الى ما يسمى إختبار الفرضيات المتعلقة باختبار دلالة الفروق بين متوسطي العينتين.

ان اختبار الاحصائي المناسب لاختبار دلالة الفروق بين وسطي المجموعتين يعتمد على حجم العينتين من جهة وعلى مدى توفر معلومات عن الانحرافات المعيارية للمجتمعين الاحصائين، بحيث

يســتخدم الاحصـائي (Z) عندما يكون حجم العينتين كـبيراً نسبياً (ن ≥ 30) وفي حالة معرفة الانحرافات المعيارية للمجتمعين وذلك وفق المعادلة التالية :

$$ز = \frac{\bar{س}_1 - \bar{س}_2}{\sqrt{\frac{ع_1^2}{ن_1} + \frac{ع_2^2}{ن_2}}}$$

حيث $\bar{س}_1$: المتوسط الحسابي للعينة الاولى .

$\bar{س}_2$: المتوسط الحسابي للعينة الثانية

$ع_1^2$: التباين للمجتمع الاول

$ع_2^2$: التباين للمجتمع الثاني

$ن_1$: حجم العينة الاولى

$ن_2$: حجم العينة الثانية

وتصاغ فرضيات الدراسة على النحو التالي :

$H_0: \mu_1 \neq \mu_2 : \mu_1 - \mu_2 \neq 0$

$H_1: \mu_1 = \mu_2 : \mu_1 - \mu_2 = 0$

أو $\mu_1 < \mu_2 : \mu_1 - \mu_2 < 0$

أو $\mu_1 > \mu_2 : \mu_1 - \mu_2 > 0$

أما في حالة عدم توفر معلومات عن الانحرافات المعيارية لمجتمعي الدراسة ؛ أي عندما تكون مثل هذه الانحرافات مجهولة ، فيمكن الاستعاضة عن الانحرافات المعيارية للمجتمعين واستخدام الانحرافات المعيارية للعينتين بدلاً منهما شريطة ان يكون حجم العينتين كبيراً نسبياً . ولكن في حالة كون حجم العينتين صغيراً فعندها يتم استخدام الاحصائي (T) بدلاً من الاحصائي (Z) ويستعاض عن الانحرافات المعيارية للمجتمع باستخدام الانحرافات المعيارية للعينتين وذلك وفق المعادلة التالية :

$$ت = \frac{\bar{س}_1 - \bar{س}_2}{\sqrt{ع^2\left(\frac{1}{ن_1} + \frac{1}{ن_2}\right)}}$$

$$ع^2 = \frac{(ن_1 - 1)\,ع_1^2 + (ن_2 - 1)\,ع_2^2}{ن_1 + ن_2 - 2}$$

حيث $ع^2$: التباين المشترك.

$ع_1^2$: تباين العينة الاولى

$ع_2^2$: تباين العينة الثانية.

$ن_1$: حجم العينة الاولى

$ن_2$: حجم العينة الثانية

$ن_1 + ن_2 - 2$: درجات الحرية.

مثال (9:8) : قام باحث بأخذ عينة من الذكور عدد افرادها (20) فرداً واخرى من الاناث عدد افرادها(15) انثى، وقاس ذكاء الافراد في العينتين، فكان المتوسط الحسابي لذكاء الذكور (110) بانحراف معياري مقداره (14)، في حين كان المتوسط الحسابي لذكاء الاناث (105) بانحراف معياري مقداره(13). عند مستوى الدلالة $\alpha = 0.05$ اختبر دلالة الفرق بين متوسطي ذكاء الذكور والاناث.

الحل:

$$H_0 : \mu_1 = \mu_2 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 : \mu_2 - \mu_2 \neq 0$$

$$ت = \frac{\bar{س}_1 - \bar{س}_2}{\sqrt{ع^2\left(\frac{1}{ن_1} + \frac{1}{ن_2}\right)}}$$

$$ع^2 = \frac{(15-1)(12)^2 + (20-1)(14)^2}{20 + 15 - 2}$$

$$\frac{19\times 196 + 14 \times 144}{35 -2} = \frac{3724 + 2016}{33}$$

$$\frac{5740}{33} = 173.9$$

$$ت = \frac{110-105}{\sqrt{173.9\left(\frac{1}{20} + \frac{1}{15}\right)}}$$

وعليه فإن

$$= \frac{5}{\sqrt{173.9 \times (0.05 + 0.067)}}$$

$$= \frac{5}{\sqrt{20.29}} = \frac{5}{4.50} = 1.11$$

منطقة الرفض (قيمة ت) الحرجة عند مستوى الدلالة $\frac{\alpha}{2} = \frac{0.05}{2}$

ودرجات حرية (20+15-2) هي : 2.021

القرار: نرفض الفرضية الصفرية إذا كان |ت| المحسوبة اكبر من قيمة(ت) الحرجة . ولما كانت قيمة (ت) المحسوبة تساوي 1.11 وهي اقل من قيمة (ت) الحرجة (2.021)، فالقرار هو عدم رفض الفرضية الصفرية؛ اي ان البيانات المتوفرة لدينا ترى عدم وجود اختلاف بين متوسط ذكاء الذكور ومتوسط ذكاء الإناث .

خامساً: اختيار دلالة الفرق بين متوسطين لعينتين مترابطتين: في بعض الحالات تكون العينتين مترابطتين بشكل أو بآخر، بحيث ان البيانات فيها تتداخل معاً أو تؤثر في بعضها البعض .

فعلى سبيل المثال قد تجد أن درجات مجموعة من الطلاب في مادة الرياضيات ترتبط الى درجة ما بدرجاتهم في مادة العلوم، اذ يمكن التنبؤ بدرجاتهم في مادة العلوم من خلال معرفة درجاتهم في مادة

الرياضيات . ففي مثل هذه الحالة وغيرها يتوفر لدى المجموعة عينتين من البيانات احدها تمثل درجات مادة العلوم والأخرى تمثل درجات مادة الرياضيات ، مثل هذه البيانات تكون مرتبة في ازواج كما هو موضح في الجدول التالي :

الطالب	1	2	3	4	5	6	7	8	9	10	11	12
درجات العلوم	75	63	57	80	69	67	95	43	52	93	70	56
درجات الرياضيات	72	58	50	82	65	73	89	46	55	91	74	50

فكما يلاحظ في الجدول اعلاه ، ان لكل طالب زوج من العلامات ، احدهما لمادة العلوم والاخرى لمادة الرياضيات ، وهذا يعني ان لدى هذه المجموعة من التلاميذ عينتين مترابطتين ، احدهما درجات مادة العلوم والاخرى درجات مادة الرياضيات .

في مثل هذه الحالات وغيرها يتطلب الامر اختبار دلالة الفرق بين متوسطي المجموعة الواحدة على العينتين المترابطتين وذلك للوقوف على مدى اختلاف هذين المتوسطين عن بعضهما البعض ، بحيث يتم استخدام الاحصائي (ت) لحساب قيمة (ت) للفرق بين المتوسطين من خلال المعادلة التالية :

$$ت = \frac{\bar{ف}}{ع_{ف}\sqrt{ن}}$$

حيث $\bar{ف}$: هو المتوسط الحسابي للفرق بين ازواج البيانات في العينتين المترابطتين .

$ع_{ف}$: الانحراف المعياري للفرق بين ازواج البيانات

ن : عدد افراد العينة .

مثال (9:9) : في دراسة قام بها باحث على مجموعة من الطلاب عدد افرادها (20) طالباً كان يهدف من خلالها المقارنة بين متوسط تحصيلهم في مادة اللغة العربية وتحصيلهم في مادة التربية الاسلامية ، وذلك لتحديد ما إذا كان تحصيلهم في هاتين المادتين مستقلاً عن بعضهما البعض ، فإذا كان المتوسط الحسابي لتحصيلهم في مادة التربية الاسلامية (68) ومتوسط تحصيلهم في مادة التربية الاسلامية (65) وكان المتوسط الحسابي للفروق بين درجاتهم على مادة التربية الاسلامية والاجتماعيات يساوي (12) بانحراف معياري مقداره (1.5) . اختبر دلالة الفروق بين متوسط تحصيل الطلاب في مادة التربية الاسلامية ومادة الاجتماعيات عند مستوى الدلالة $\alpha = 0.01$.

الحل:

$$H_0: \mu_1 = \mu_2 : \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 : \mu_1 - \mu_2 \neq 0$$

$$ت = \frac{\bar{ف}}{ع_{ف}\sqrt{ن}}$$

$$\frac{12}{4.47 \times 1.5} = \frac{12}{6.71} = 1.79$$

منطقة الرفض (قيمة ت) الحرجة عند مستوى الدلالة $\frac{\alpha}{2} = \frac{0.01}{2}$

ودرجات حرية (ن–1 : 20-1) تساوي 2.093.

القرار: نرفض الفرضية الصفرية اذا كان |ت| المحسوبة اكبر من قيمة (ت) الحرجة. ، لما كانت قيمة |ت| المحسوبة أقل من قيمة (ت) الحرجة؛ اي $1.79 < 2.093$، فالقرار اذا هو عدم رفض الفرضية الصفرية؛ وهذا يعني ان تحصيلهم في مادة التربية الاسلامية غير مستقل عن تحصيلهم في مادة الاجتماعيات.

سادساً: اختيار الفرق بين نسبتين

كما هو الحال في اختبار الفرق بين متوسطين لعينتين مستقلتين، يمكن اختبار الفرق بين نسبتين لعينتين يتم اخذهما من مجتمعين منفصلين أو من مجتمع واحد.

ففي بعض الحالات يقتضي الامر المقارنة بين عينتين او مجتمعين من حيث نسبة الحالات التي توجد في كل منهما. فعلى سبيل المثال، قد يتطلب الامر مقارنة بين نسبة الذكور ونسبة الاناث اللواتي يؤيدون عمل المرأة أو المقارنة بين نسبة اعضاء هيئة التدريس في جامعة ما ممن يؤيدون التدريس باستخدام الحاسوب ونسبة الذين لا يؤيدون. ففي مثل هذه الحالات وغيرها يتم اختبار دلالة الفروق بين النسب او الاستدلال على النسب في المجتمعات من خلال النسب التي توفرها العينات، بحيث يتم استخدام المعادلة التالية لتحديد القيم الاحصائية التي تستخدم في المقارنة:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

حيث $\hat{P}_1$: نسبة العناصر في العينة الاولى

$\hat{P}_2$: نسبة العناصر في العينة الثانية

$(\hat{P}_1 - 1): P_1$

$(\hat{P}_2 - 1): P_2$

n_1 : حجم العينة الاولى

n_2 : حجم العينة الثانية

وتصاغ الفرضيات في اختيار دلالة الفرق بين النسب على النحو الآتي :

$H_0 : P_1 = P_2 : P_1 - P_2 = 0$

$H_1 : P_1 \neq P_2 : P_1 - P_2 \neq 0$

أو $P_1 < P_2 : P_1 - P_2 < 0$

$P_1 > P_2 : P_1 - P_2 > 0$

مثال (9:10) : اراد باحث أن يختبر الفرق بين نسبة الطلاب الذين يؤيدون فكرة استخدام نظام الساعات المعتمدة في المدارس من كلا الجنسين . قام باختيارعينة عشوائية من الذكور عدد افرادها (500 طالب) ووجد ان (350) طالباً منهم يؤيدون فكرة الساعات المعتمدة، كما اختيار عينة اخرى من الاناث عدد افرادها (400 طالبة) ووجد ان (300) طالبه تؤيد فكرة الساعات المعتمدة. اعتماداً على البيانات اعلاه اختبر دلالة الفرق بين نسبة الذكور الذين يؤيدون نظام الساعات المعتمدة ونسبة الاناث اللواتي يؤيدون نظام الساعات المعتمدة عند مستوى الدلالة $\alpha = 0.05$.

الحل:

$H_0 : P_1 = P_2 : P_1 - P_2 = 0$

$H_1 : P_1 \neq P_2 : P_1 - P_2 \neq 0$

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

$$0.70 = \frac{350}{500} = \frac{X_1}{n_1} = \hat{P}_1$$

$$0.75 = \frac{300}{400} = \frac{X_2}{n_2} = \hat{P}_2$$

$$0.30 = (0.70 - 1) = (\hat{P}_1 - 1): P_1$$

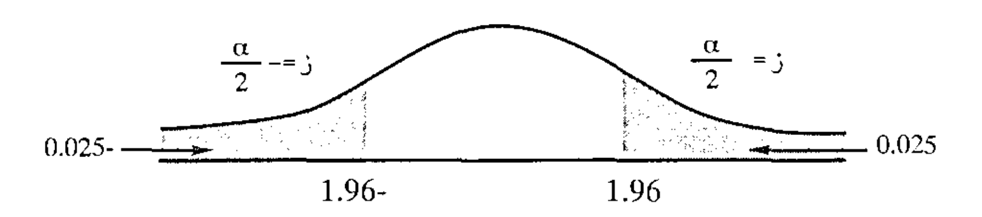
$$Z = \frac{0.70 - 0.75}{\sqrt{\frac{0.30(1-0.30)}{500} + \frac{0.25(1-0.25)}{400}}}$$

$$= \frac{-0.30}{\sqrt{\frac{0.30 \times 0.70}{500} + \frac{0.25 \times 0.75}{400}}}$$

$$= \frac{-0.30}{\sqrt{\frac{0.21}{500} + \frac{0.1875}{400}}}$$

$$= \frac{-0.30}{\sqrt{0.00042 + 0.00047}} = \frac{-0.30}{0.030} = -10$$

منطقة الرفض: هي قيمة (Z) عند مستوى الدلالة $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ وهي تساوي (1.96)



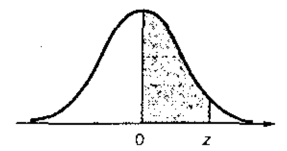
القرار : نرفض الفرضية الصفرية اذا كان مطلق قيمة (Z) المحسوبة اكبر من قيمة (Z) الحرجة . وبما ان |10 -| اكبر من قيمة (1.96)، فعليه فأن القرار هو عدم قبول الفرضية الصفرية ؛ اي ان البيانات المتوفرة لدينا لا تدعم تساوي نسبة الذكور الذين يؤيدون نظام الساعات المعتمدة ونسبة الاناث اللواتي يؤيدن نظام الساعات المعتمدة . وهذا يعني أن هناك اختلافاً بين نسبة الذكور التي تؤيد فكرة استخدام الساعات المعتمدة ونسبة الإناث اللواتي يؤيدن ذلك .

ملحق رقم (1)
جدول الأرقام العشوائية

87024	74221	69721	44518	58804	04860	18127	16855	61558	15430
04852	03436	72753	99836	37513	91341	53517	92094	54386	44563
33592	45845	52015	72030	23071	92933	84219	39455	57792	14216
68121	53688	56812	34869	28573	51079	94677	23993	88241	97735
25062	10428	43930	69033	73395	83469	25990	12971	73728	03856
78183	44396	11064	92153	96293	00825	21079	78337	19739	13684
70209	23316	32828	00927	61841	64754	91125	01206	06691	50868
94342	91040	94035	02650	36284	91162	07950	36178	42536	49869
92503	29854	24116	61149	49266	82303	54924	58251	23928	20703
71646	57503	82416	22657	72359	30085	13037	39608	77439	49318
51809	70780	41544	27828	84321	07714	25865	97896	01924	62028
88504	21620	07292	71021	80929	45042	08703	45894	24521	49942
33186	49273	87542	41086	29615	81101	43707	87031	36101	15137
40068	35043	05280	62921	30122	65119	40512	26855	40842	83244
76401	68461	20711	12007	19209	28259	49820	76415	51534	63574
47014	93729	74235	47808	52473	03145	92563	05837	70023	33169
67147	48017	90741	53647	55007	36607	29360	83163	79024	26155
86987	62924	93157	70947	07336	49541	81386	26968	38311	99805
58973	47026	78574	08804	22960	32850	67944	92303	61216	72948
71635	86749	40369	94639	40731	54012	03972	98581	45604	34865
60971	54212	32596	03052	84150	36798	62635	26210	95685	87009
06599	60910	66315	96690	19039	39878	44688	65146	02482	73130
89960	27162	66264	71024	18708	77974	40473	87155	35834	03114
03930	56898	61900	44036	90012	17673	54167	82396	39468	49566
31338	28729	02095	07429	35718	86882	37513	51560	08872	33717
29782	33287	27400	42915	49914	68221	56088	06112	95481	30094
68493	88796	94771	89418	62045	40681	15941	05962	44378	64349
42534	31925	94158	90197	62874	53659	33433	48610	14698	54761
76126	41049	43363	52461	00552	93352	58497	16347	87145	73668
80434	73037	69008	36801	25520	14161	32300	04187	80668	07499
81301	39731	53857	19690	39998	49829	12399	70867	44498	17385
54521	42350	82908	51212	70208	39891	64871	67448	42988	32600
82530	22869	87276	06678	36873	61198	87748	07531	29592	39612
81338	64309	45798	42954	95565	02789	83017	82936	67117	17709
58264	60374	32610	17879	96900	68029	06993	84288	35401	56317
77023	46829	21332	77383	15547	29332	77698	89878	20489	71800
29750	59902	78110	59018	87548	10225	15774	70778	56086	08117
08288	38411	69886	64918	29055	87607	37452	38174	31431	46173
93908	94810	22057	94240	89918	16561	92716	66461	22337	64718
06341	25883	42574	80202	57287	95120	69332	19036	43326	98697
23240	94741	55622	79479	34606	51079	09476	10695	49618	63037
96370	19171	40441	05002	33165	28693	45027	73791	23047	32976
97050	16194	61095	26533	81738	77032	60551	31605	95212	81078
40833	12169	10712	78345	48236	45086	61654	94929	69169	70561
95676	13582	25664	60838	88071	50052	63188	50346	65618	17517
28030	14185	13226	99566	45483	10079	22945	23903	11695	10694
60202	32586	87466	83357	95516	31258	66309	40615	30572	60842
46530	48755	02308	79508	53422	50805	08896	06963	93922	99423
53151	95839	01745	46462	81463	28669	60179	17880	75875	34562
80272	64398	88249	06792	98424	66842	49129	98939	34173	49883

ملحق رقم (2)

جدول التوزيع الطبيعي (z)

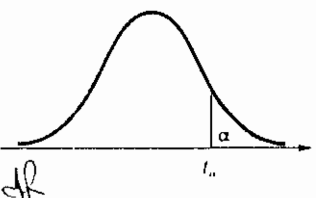


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas*. New York: John Wiley & Sons, Inc., 1952. Reproduced by permission of A. Hald and the publisher.

ملحق رقم (3)

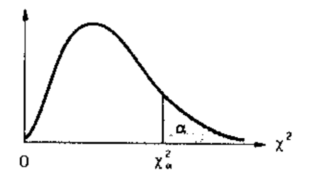
جدول توزيع ت t- students



ν	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source. From E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., *Biometrika*, 1966. Reproduced by permission of the *Biometrika* Trustees.

ملحق رقم (٤)
جدول توزيع كاي² (X^2)



DEGREES OF FREEDOM	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$
1	.0000393	.0001571	.0009821	.0039321	.0157908
2	.0100251	.0201007	.0506356	.102587	.210720
3	.0717212	.114832	.215795	.351846	.584375
4	.206990	.297110	.484419	.710721	1.063623
5	.411740	.554300	.831211	1.145476	1.61031
6	.675727	.872085	1.237347	1.63539	2.20413
7	.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581
150	109.142	112.668	117.985	122.692	128.275
200	152.241	156.432	162.728	168.279	174.835
300	240.663	245.972	253.912	260.878	269.068
400	330.903	337.155	346.482	354.641	364.207
500	422.303	429.388	439.936	449.147	459.926

DEGREES OF FREEDOM	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	9.48773	11.1433	13.2767	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5346	20.0902	21.9550
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9871	18.3070	20.4831	23.2093	25.1882
11	17.2750	19.6751	21.9200	24.7250	26.7569
12	18.5494	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3621	24.7356	27.6883	29.8194
14	21.0642	23.6848	26.1190	29.1413	31.3193
15	22.3072	24.9958	27.4884	30.5779	32.8013
16	23.5418	26.2962	28.8454	31.9999	34.2672
17	24.7690	27.5871	30.1910	33.4087	35.7185
18	25.9894	28.8693	31.5264	34.8053	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5822
20	28.4120	31.4104	34.1696	37.5662	39.9968
21	29.6151	32.6705	35.4789	38.9321	41.4010
22	30.8133	33.9244	36.7807	40.2894	42.7956
23	32.0069	35.1725	38.0757	41.6384	44.1813
24	33.1963	36.4151	39.3641	42.9798	45.5585
25	34.3816	37.6525	40.6465	44.3141	46.9278
26	35.5631	38.8852	41.9232	45.6417	48.2899
27	36.7412	40.1133	43.1944	46.9630	49.6449
28	37.9159	41.3372	44.4607	48.2782	50.9933
29	39.0875	42.5569	45.7222	49.5879	52.3356
30	40.2560	43.7729	46.9792	50.8922	53.6720
40	51.8050	55.7585	59.3417	63.6907	66.7659
50	63.1671	67.5048	71.4202	76.1539	79.4900
60	74.3970	79.0819	83.2976	88.3794	91.9517
70	85.5271	90.5312	95.0231	100.425	104.215
80	96.5782	101.879	106.629	112.329	116.321
90	107.565	113.145	118.136	124.116	128.299
100	118.498	124.342	129.561	135.807	140.169
150	172.581	179.581	185.800	193.208	198.360
200	226.021	233.994	241.058	249.445	255.264
300	331.789	341.395	349.874	359.906	366.844
400	436.649	447.632	457.305	468.724	476.606
500	540.930	553.127	563.852	576.493	585.207

الإحصاء التربوي



يشهد هذا العصر الذي نعيشه تغيراً واضحاً وملموساً في كافة ميادين الحياة ولا سيما أننا نعيش طفرة الانفتاح والتفجير المعرفي والتفوق التكنولوجي والتي زادت من حجم المطالب والأعباء الإنسانية . إن مثل هذه التغيرات المتسارعة ، مما لا يدعو للشك أصبحت تتطلب التخطيط الحذر والمستمر للتعامل معها ومواجهة المستقبل . ومن هنا ، نجد أن الاهتمام في استخدام المناهج والأساليب الإحصائية المختلفة في توظيفها في دراسة وتحليل الخصائص المرتبطة بمختلف أنواع المتغيرات والعوامل بات يشكل عنصراً هاماً في كافة الميادين : الاجتماعية والنفسية والتربوية والصناعية والاقتصادية والعلمية وغيرها من الميادين الأخرى .

وبالرغم من تباين الأساليب الاحصائية وتعدد أغراضها واستخداماتها ، إلا أنها في مجموعها تساعد في وصف البيانات الكبيرة المتعلقة بالتغيرات الإحصائية وفي عملية تحليلها ، الأمر الذي يساعد في عمل الإستنتاجات وإصدار الأحكام واتخاذ القرارات المناسبة التي تساعد في عمليات التخطيط .



دار الشروق للنشر والتوزيع

المركز الرئيسي : عمان / الأردن - تلفون 4618190-4618191 - 4624321

فاكس 4610065 - ص.ب 926463 - عمان 11118 الأردن

فرع الجامعة الأردنية - تلفون : 5358352

E-mail: shorokjo@nol.com.jo

ww.shorok.com

وكلاؤنا في فلس

دار الشروق للنشر والتوزيع - رام الله - المنارة - تلف

دار الشروق للنشر والتوزيع - غزة - الرمال الجنوبي - ت

المراجع

(1) البلداوي ، عبد الحميد البلداوي (2004) ، أساليب البحث العلمي والتحليل الاحصائى : - التخطيط للبحث وجمع وتحليل البيانات يدوياً وباستخدام برنام spss، دار الشروق.

(2) عدس، عبد الرحمن؛ والمينزل عبد الله (2004) ، مقدمة في الاحصاء التربوي، دار الفكر، عمان- الأردن.

(3) عودة، أحمد (1993) القياس والتقويم في العملية التدريسية ، دار الملى، اربد ، الأردن.

(4) ماضي ، محمد طاهر؛ وعثمان، ماجد إبراهيم (1999) ، الاحصاء في التربية وعلم النفس ، دار القلم للنشر والتوزيع، الإمارات العربية المتحدة.

(5) النبهان، موسى (2001) أساسيات الاحصاء في التربية والعلوم الإنسانية والاجتماعية. مكتبة الفلاح للنشر والتوزيع. العين ، الامارات العربية المتحدة.

Christetens, L.B & Stoup, C.M. (1986). Introduction to statistics For(6) the social and behavioral sciences . Brooks Cole Publishing company.

Draper, N., & Smith, H. (1990) Applied regession analysis. John Wiley(7) & Sons Inc. London.

Hieman, G. (1992). Basic statistics For the behavioral sciences.(8) Houghton Miffin Company.

Kiess, H.O. & Bloomquist, D.W. (1985) Psychological research(9) methods: Aa conceptual approach. Allyn & Bacon, Inc.

Sineich, T. (1993) Statistics by example (5th^{ed}). Macmillan publishing(10) company.